

## Local Unit Alignment, Reorganization, and Consolidation Commission

# Literature Review and Analysis Related to Measurement of Local Government Efficiency

Marc Holzer, Ph.D.  
Principal Investigator

John Fry  
Project Director

Etienne Charbonneau  
Deputy Project Director

Norma Riccucci, Ph.D.  
Supervising Faculty

Etienne Charbonneau  
Sunjoo Kwak  
Senior Research Associates

Eileen Burnash  
Librarian



May 6, 2009

## **TABLE OF CONTENTS**

# **Literature Review and Analysis Related to Measurement of Local Government Efficiency**

<b>EXECUTIVE SUMMARY</b>	<b>1</b>
<b>PURPOSE OF THE REPORT</b>	<b>6</b>
<b>REVIEW AND SYNTHESIS OF THE LITERATURE</b>	<b>9</b>
<b>COMMENTS ON THE LITERATURE</b>	<b>23</b>
<b>CONCLUSIONS</b>	<b>24</b>
<b>APPENDIX A – ANNOTATED BIBLIOGRAPHY</b>	<b>26</b>
<b>APPENDIX B – METHODOLOGY</b>	<b>47</b>
<b>APPENDIX C – RFP QUESTIONS</b>	<b>51</b>
<b>APPENDIX D – BIBLIOGRAPHIC DATABASES</b>	<b>52</b>

## EXECUTIVE SUMMARY

The Local Unit Alignment, Reorganization, and Consolidation Commission (LUARCC) contracted with the School of Public Affairs and Administration (SPAA) at Rutgers University, Newark campus in October of 2008 to conduct a review of the literature on *measurement of local government efficiency*. Although centralized performance measurement systems have been in place and continue to evolve and improve in other countries, they are absent at a state level in the United States. The New Jersey legislature has recognized the value of such a system in the enabling legislation creating LUARCC:

Local governments must be trained to use performance measures for decision making, strategic planning, performance improvement, accountability, and communication, and rewarded for increased efficiencies that result from their use.

This report provides an overview of performance measurement systems in the United States and throughout the world. Based on the components found in the systems the literature describes, a performance measurement system should include the following:

- A system of benchmarking among the local governments in the state in order to facilitate improvement and best practices.
- A system of different types of performance measures including workload, input, output, and outcome indicators.
- A system of management review that encourages management action and monitors its effectiveness.
- Reporting to the state of performance measures and management initiatives resulting from the analysis and review of the performance measures.

According to the Public Performance Measurement and Reporting Network (2009), performance measurement and reporting serves three purposes:

- Informs government policy and management practices and should lead to: performance improvement and policy development that results in better practice and alignment of governmental services to community needs.
- Provides data that enables accountability and transparency.
- Enables informed communication between citizens, government, and not-for-profit organizations to foster trust.

Although it is a complex undertaking, the most difficult decision for LUARCC and the state is not what the overall system should look like when it has matured. The most difficult decision is how to get there from the current situation. Our review of the published research suggests that the following five principles should guide the development of a measurement system:

- Measurement should occur at each stage of the service delivery process.
- Measurement should start with simple, basic forms of data analysis, eventually progressing to the powerful tool of Data Envelopment Analysis (DEA).
- The system should reflect “best practice” models for performance reviews and improvements, such as CitiStat or CompStat.
- Municipalities should benchmark their performance against comparable cities and towns, and should collect standardized data to do so.
- Performance data should be available to the public.

Overall, the literature addressed some questions in the scope of work for this report:

- There are many models and techniques for measuring efficiency.
- Basic information can now be collected, but valid benchmarking requires that municipal accounting and data collection systems be standardized.
- Stakeholders should use the data for diagnosis of efficiencies and improvement of services.
- Best practices in performance measurement are illustrated by the “Stat” cases: CitiStat, CompStat, etc.
- Non-financial benefits of performance measurement include improvements in the quality and efficacy of services.

However, some questions in the scope of work are not addressed:

- Municipal performance is more commonly compared to past performance than to the level of services in another jurisdiction.
- When the literature estimates cost savings, it does not generally make the link to tax savings. In one instance, a series of newspaper articles revealed this was a policy choice being debated by the governing body.

- Government structure is not related to measurement of efficiency.
- Other states have not mandated performance measurement by municipalities.

**Questions from the Scope of Work (RFP) on Measuring Local Government Efficiency**

Purpose of the Literature Review:

Through analysis of existing case studies, this work examines attempts to measure productivity outcomes through the application of a methodology which is credible according to the investigator. The primary purpose is to ascertain whether there are methodologies of benchmarking, performance evaluation or “best practices” for determining the most efficient delivery of municipal services within the United States.

**How was efficiency measured?**

*The commonly used "cost per capita" measure of efficiency has issues of comparability and usefulness. The literature provides a number of different indicator types used in a performance measurement system in order to provide useful data to managers and others. The body of the report contains references to other resources, which list specific indicators.*

**Is there any correlation between size of service area for the delivery of a given service and efficiency and/or effectiveness?**

*The report on Optimal Size covers this topic, and indicates that much of the literature states that relatively large and relatively small municipalities may be less efficient, but this varies with the specific service being delivered. There is some disagreement in the literature about optimal size even within a specific service type (police services are an important example).*

**What type of information was collected, and any changes that needed to be made in municipal record keeping to report the data consistently over time?**

*If users or stakeholders are to make inter-jurisdictional comparisons, then measurements must be standardized. The literature states that it is easier to measure efficiency if comparisons are only historical within the same unit, but such comparisons have limited usefulness. Measurement of outcomes requires changes in record keeping, since the focus, where performance has been measured, is typically on workload or output measures.*

**How was the information utilized?**

*Data was used primarily for internal management, and less frequently for public reporting and accountability.*

**Was the municipality compared to others in a larger geographic study or was the municipality compared only to its own progress over time?**

*The most common comparisons are to prior time periods within the same government unit, but these comparisons are not as useful as inter-jurisdictional comparisons. The report provides examples of systems in which inter-jurisdictional comparisons have been made.*

**Did evaluative studies identify cost savings which resulted in property tax savings through greater attention to municipal productivity?**

*The literature rarely links examples of actual cost savings to tax savings. Cost savings, not tax savings, are documented, but not that frequently and with suspicions about reliability.*

**Were there secondary benefits to the measurement of productivity that were identified in the case studies other than cost savings?**

*The literature promotes quality of service delivery as a major, perhaps even primary, benefit of performance measurement.*

**Were benchmarks established and if so, how was the benchmark utilized?**

*The literature defines benchmarks primarily as inter-jurisdictional comparisons. But the most common form of comparison is over time within the same unit of government. Benchmarks were utilized as catalysts to performance improvement. An example from the North Carolina Local Government Performance Measurement Project is the use of inter-jurisdictional comparisons on solid waste efficiency to change the process for garbage collection.*

**Were changes in government structure or other government reforms effectuated as a result of productivity benchmarking?**

*Changes in governmental structure are not typically a direct result of benchmarking. There are instances when the identification of inefficient service delivery leads to changes, which could include a change in structure. Performance measurement encourages changes that are usually operational in nature.*

**Were best practices established and if so, how were they established and how were they used?**

*While there are a number of performance measurement models that could be considered as best practices, the literature did not highlight the issue of finding best practices in service delivery as a result of the measurement of performance.*

**Does the literature point to a model of efficiency measurement we might use in New Jersey to improve municipal service delivery performance?**

*There are models based on inter-jurisdictional consortiums, such as those developed by ICMA and the University of North Carolina Institute of Government. The “Stat” models are more oriented to performance management for results or outcomes. The State of New Jersey could examine these and other models to determine what would fulfill the performance measurement goals of the State.*

**What role did the State have, if any, in other States where serious attempts were made at measuring service delivery or was the impetus only from within the municipality itself?**

*There is no state-mandated model or performance measurement system. In some other countries, centralized systems have been developed and implemented by a central government to require local governments to provide the data.*

## PURPOSE OF THE REPORT

To meet its overall goals as expressed in the legislation signed into law on March 15, 2007, the Local Unit Alignment, Reorganization, and Consolidation Commission (LUARCC) commissioned a literature review to examine best practices among municipalities in the United States and elsewhere in measuring service delivery performance. In the RFP it issued, LUARCC elaborated that the review will examine the attempts to measure productivity outcomes and will ascertain whether there are methodologies of benchmarking, performance evaluation, or best practices for determining the most efficient delivery of municipal services within the United States. The RFP focused on the following points:

- How is efficiency measured?
- Is there any correlation between size of service area for the delivery of a given service and efficiency and/or effectiveness? (Note that the report on Optimal Municipal Size and Efficiency is devoted to this topic.)
- What type of information was collected? What changes need to be made in municipal record keeping to report the data consistently over time?
- How was the information utilized?
- Was the municipality compared to others in a larger geographic study or was the municipality compared only to its own progress over time?
- Did evaluative studies identify cost savings which resulted in property tax savings through greater attention to municipal productivity?
- Were there secondary benefits to the measurement of productivity that were identified in the case studies other than cost savings?
- Were benchmarks established and if so, how was the benchmark utilized?
- Were changes in government structure or other government reforms effectuated as a result of productivity benchmarking?
- Were best practices established and if so, how were they established and how were they used?
- Does the literature point to a model of efficiency measurement we might use in New Jersey to improve municipal service delivery performance?
- What role did the State have, if any, in other States where serious attempts were made at measuring service delivery or was the impetus only from within the municipality itself?

The annotated bibliography that is included in this report is the result of the research into the questions above. SPAA has synthesized the information in the articles that are listed in the bibliography and highlighted the conclusions presented in this report. Our role was to look for model implementations of performance measurement systems and provide evidence of what must be put in place to utilize them effectively. The enabling legislation for the establishment of LUARCC states:

Local governments must be trained to use performance measures for decision making, strategic planning, performance improvement,

accountability, and communication, and rewarded for increased efficiencies that result from their use.

The literature on performance measurement in local government is vast. “From New York City to government entities of all shapes and sizes, performance measurement continues to gain in popularity. Robert Kaplan's Balanced Scorecard, David Osborne's Reinventing Government, the success of the NYPD's COMPSTAT process, and the ever-increasing public pressure to report performance have made the practice of performance measurement commonplace in the public sector.” (Page and Malinowski, 2004)

LUARCC has asked SPAA at Rutgers to review the literature to provide insight into how to measure the efficient delivery of municipal services. Our review describes types of efficiency measures, surveys some of the techniques for analysis of the resulting data, discusses the important methods of targeting and comparing results, and highlights some of the benefits resulting from performance measurement. The final section attempts to describe a performance measurement model for New Jersey based on what exists in the literature, and, although the review highlights the best practices in the field, this discussion recognizes that getting to that level will require time and a process of educating and training local government officials and staff.

## **A Primer on Performance Measurement**

Since this field has evolved so quickly in recent decades and can be quite complex, it is useful to introduce some terminology and an overall model of performance measurement. Over the years there have been many categorizations of performance measures and there has been accompanying disagreement about what the categories should be and what is really important. For the purposes of this review, the following types of indicators provide the framework of the discussions that follow (Julnes and Holzer, 2008; Fry, 2004):

- Workload – indicators of the amount of work that must be accomplished, such as the number of patients requiring treatment or the curb miles that are required to be swept.
- Inputs – indicators of the resources used, such as budgeted dollars or FTE's.
- Outputs – the work completed, such as patients treated or curb miles swept.
- Exogenous or explanatory variables – factors that affect the work process but are not a part of it, such as the education level of the population served or the number of days without snow cover.
- Outcomes – the results of government service delivery, such as infant mortality or cleanliness of the streets.

A simple model of service delivery views the workload as the amount of work to be done, which is a variable determined by the conditions or the population served. The inputs are determined by government executives or high-level managers. Outputs are affected by these input decisions, as well as by the effectiveness of the operational management and the productivity of the employees. Outcomes are the ultimate result of

all the other indicator types, including factors over which government does not have direct control, such as workload measures and exogenous variables.

The following is a theoretical discussion. In a more sophisticated model of service delivery, the analysts realize output may be almost completely determined by workloads as long as inputs are adequate. Effective management sees it as the duty of government to address workloads, and, therefore, outputs are just a reflection of having done the job. Enlightened government may look for ways to reduce the workloads, investing additional inputs into programs targeted to reducing the need for service.

The following concrete example is from Project Scorecard (The United States Conference of Mayors, 2000), a street and sidewalk litter rating system begun in the 1970's in New York City and now part of the Mayor's Management Report. In a given sanitation district, the workload would be the curb miles to be swept and the frequency of sweeping that was required. The input would be the sweeper hours (which results from the general budget inputs of manpower and equipment). The output would be the curb miles swept. Exogenous variables could include the number of days with winds over ten miles per hour, the percent of commercial land use (litter creating), or the education level of the population. The outcome was the cleanliness of the streets as measured by the city or its contractor through a trained observer rating of cleanliness on samples of streets in the sanitation district.

The more sophisticated model recognizes that the real workload is the amount of litter generated on the streets, which drives the frequency of sweeping that is necessary. If the workload increases and resources are adequate to the task, the output of curb miles swept will increase, but the outcome of street cleanliness may not be affected. Enlightened government will look for ways to reduce the workload, that is, the amount of litter generated. In New York City, they used Sanitation Police to reduce the workload by writing tickets for improper containerization of garbage (reducing spillage onto the street and sidewalk), for littering, and for alternate side of the street parking (to increase the effectiveness of the sweepers). One of the results of the enlightened evaluation of the Scorecard outcome measure of street and sidewalk cleanliness was the civilianization of the Sanitation Police, which allowed more enforcement effort at less labor cost, a direct attack on workload.

The literature discusses indicator types in addition to the five presented. The basic five indicator types yield efficiency, productivity, and effectiveness measures from simple calculations. Among the many sources for examples of specific indicators is SPAA's Public Performance Measurement and Reporting Network website ([www.ppmrn.net](http://www.ppmrn.net)). It is a guide to other websites and manuals. Many of these will contain lists of indicators and other information. Many of the references in the annotated bibliography list other indicators. (Braadbaart, 2007; Garcia-Sanchez, 2006; New South Wales Department of Local Government, 2008; Drake and Simper, 2000; Sun, 2002)

## REVIEW AND SYNTHESIS OF THE LITERATURE

### Efficiency Measures

Measurement of efficiency is not an easy task. Attempts to do so have been going on since at least the 1920's (Ridley, 1927), but the growth in the literature over the last fifteen years (Government Accounting Standards Board, 1994; International City/County Management Association, 2008) is a testimony to the overwhelming increase in the interest in measuring performance and, it appears, to the belief it can promote improvement. The experts, who have toiled to get it right, know the difficulties, as well as the potential:

In the face of these complexities, too many local governments resort to reporting "FTEs per 1,000 population" or "cost per capita" for services overall or for the services of a particular department. These are extremely crude measures of efficiency, if they can be called efficiency measures at all. (Ammons and Rivenbark, 2008)

Outcomes are reasons governments produce. Outcomes are the crime rate, the cleanliness of streets, the economic vitality of the community. However, outcomes are not totally under the control of municipal government, or any government for that matter. Behn, who writes a monthly management report, largely on performance measurement and managing for results, states in one of his articles:

The standard measurement mantra is: "Don't measure inputs. Don't measure processes. Don't measure activities. Don't measure outputs. Only measure outcomes." Unfortunately, in city government (indeed, in any government) this is often difficult. Sometimes it is impossible. Consequently, a CitiStat strategy may have to rely more on output data than on outcome data. (Behn, 2007)

The implication of Behn's comments is that the use of the data determines what one needs to measure. From an internal management perspective, the use of the data is to increase allocative efficiency, that is, to reallocate resources, to modify processes, and to reconfigure budgets to get the results desired. Outcomes, such as clean streets, are more important to the public. Outputs, such as curb miles swept, are more important to managers.

The discrepancy between larger goals and measures over which managers have control is more severe when one considers workload measures. Some managers use workload as a measure of productivity, but workload only describes the amount of work to be done. A classic example is using tons of garbage collected as a measure of productivity. If the tons-of-garbage indicator declines, it does not indicate the Department of Sanitation is slacking off. Missed or late pick-ups are measures of performance. Tonnage is not productivity. In fact, if recycling is effective, the workload measure of tons will decline, which indicates improved productivity of the solid waste program as a whole.

Ammons and Rivenbark (2008) comment on the use of workload indicators:

Officials taking the narrow view of accountability are less likely to venture beyond workload measures and are unlikely to try to incorporate performance measures into key management systems. For them, it seems rational and prudent to collect only the simplest measures and to divert as few resources as possible from service delivery to the measurement of performance. Given their narrow view of accountability and the minimal value of raw workload counts for management or policy decisions, they are unlikely to use performance measures meaningfully in strategic planning or management systems, performance contracts, departmental or individual work plans, performance targets, performance audits, program evaluations, service improvement strategies, cost – benefit analyses, annexation and other special studies, or budget proposals.

Practitioners and studies have often distinguished between controllable and non-controllable inputs to explain results or outcomes. Controllable factors include management systems and capacities, while factors that are often beyond the control of a local jurisdiction may include economic, environmental and demographic factors. In his description of Project Scorecard, Mayor Giuliani states:

The cleanliness of New York City is not determined by the performance of the Department of Sanitation alone. Many factors contribute to the cleanliness of streets in neighborhoods. Weather conditions and the collective behavior of pedestrians, motorists, homeowners, tenants, landlords, building superintendents, shopkeepers, commercial establishments and private carters are all major factors affecting street and sidewalk cleanliness. Scorecard is designed to measure actual litter conditions that are the product of the millions of "sanitation transactions" that take place every day. (The United States Conference of Mayors, 2000)

Project Scorecard does not measure all of these exogenous variables, but it includes some, like weather, and that provides additional credibility to the results, when managers are held accountable for performance.

Other articles reference the use of socioeconomic or other conditions outside the control of the government unit (Ouellette and Vierstraete, 2005; Stevens, 2005; Worthington and Dollery, 2002). An alternative approach is provided by incorporating the exogenous variables in a profile of the government and categorizing the governments based on the profile before comparing them (New South Wales Department of Local Government, 2008). Accounting for these exogenous variables, in one way or another, can reduce the estimated level of inefficiency and make it more palatable for managers to focus on outcomes over which they do not have complete control.

The Balanced Scorecard approach (Eagle, 2004; Ho and Chan, 2002) is a methodology that was developed to counteract the over-reliance on bottom line, profit-focused indicators in the private sector. The Balanced Scorecard used other measures that were related to performance and were part of the production process prior to the final determination of profitability. This allowed organizations to see problems before they hit the bottom line. The concept is used in the public sector to encourage a broad view of different parts of the process of producing municipal services.

Data must be comparable, particularly if the organization intends to use cross-jurisdictional comparisons. One author blames the demise of one system on self-reported data and their inconsistencies (Coe, 1999). To ensure different organizations in a benchmarking consortium provide comparable data, some form of auditing can be helpful, including even the inexpensive device of asking the manager to certify that the data is correct (New South Wales Department of Local Government, 2008).

The North Carolina Local Government Performance Measurement Project had difficulty in its early stages because of comparability problems (Rivenbark and Carter, 2000). “The lack of generally accepted criteria to compare service costs for local government has hindered benchmarking initiatives.” Led by the Institute of Government at the University of North Carolina, the project now contains a “full cost-accounting model developed to ensure that localities employ the same methodology to collect and report cost data associated with performance measures.” The author states further that the “accuracy and comparability of performance and cost data are the fundamental ingredients of a benchmarking and performance measurement project.” The Institute engages in a rigorous data cleaning process and returns drafts to the participating units for their review. Without the centralized resource represented by the Institute of Government, the comparability of the performance measures would be suspect and the entire project would suffer.

### **Analysis of the Measures**

Both reporting and service improvement can be products of performance measurement systems. “Perhaps it is axiomatic that performance measurement systems designed strictly for the former [i.e., performance reporting], especially when a premium is placed on ease of data collection, are unlikely to yield much of the latter” (Ammons and Rivenbark, 2008). This is not to say that the performance measures that municipalities are already collecting are of no use. At the present time, they are the only data available on which decisions can be made. However, looking past that short-term time-frame, the systematic collection of more meaningful data for all municipalities would drive better management decisions.

The basic steps in a performance measurement system include the following (United States Office of Personnel Management, 1974; Hatry and Wholey, 1999):

1. Begin with a simple measurement system and draft performance measures.

2. Collect the data on performance measures.
3. Analyze the data.
4. Interpret the data, using comparisons where possible.
5. Identify functions, units and procedures that are “ripe” for performance improvement.

Performance measurement is a routine and iterative process. Once the five steps are completed, the steps begin again. Since performance improvement is an on-going process, sometimes the steps do not appear so distinct. In particular, analyzing, comparing, and reviewing the data may occur simultaneously or, at the very least, in iterative phases.

The literature has recommended a number of different ways to analyze performance over the years, but one – Data Envelopment Analysis (DEA) – has become the standard, first internationally, but now with growing widespread acceptance in the United States. DEA calculates performance indicators based on multiple inputs and outputs and distinguishes between types of efficiency, such as scale, technical, and allocative efficiency (Drake and Simper, 2000; Sun, 2002; Ouellette and Vierstraete, 2005; Williams, 2005; Woodbury et al, 2003; Chalos and Cherian, 1995; Moore et al, 2005). DEA also separates out the effects of exogenous variables. Woodbury states: “... the present reliance on partial measures of performance is inadequate and should be heavily augmented by data envelopment analysis” and “the obvious analytical route seems to be an industry-wide adoption of DEA methodologies.”

According to Worthington and Dollery (2002), “The use of DEA as a technique for measuring the efficiency of government service delivery is now relatively well-established in Australia and several other advanced countries.” DEA assesses the efficiency of courts of law (Pedraja-Chaparro and Salinas-Jimenez, 1996), municipal water services (García-Sánchez, 2006), public schools (Mante and O’Brien, 2002) and municipal police services (Drake and Simper, 2002; Barros, 2007).

Indirect data can distort performance data and the estimates of efficiency resulting from them. In a study of health care, it appeared there was distortion in the estimates of inefficiency and the comparisons of hospital units because of the manipulation of this indirect data (Barretta, 2008). This study breaks the total costs into sub-categories in an analysis of the variations found in efficiency, concluding that accounting differences are masking the real rates of efficiency. It suggests removing the costs allocated from internal subunits in the hospital (indirect costs) because the hospitals manipulated them in order to increase perceived efficiency.

Dashboard analysis has also become popular. Like the automobile device of the same name, it shows indicators of performance, including an alarm status, when appropriate. Writing about the Washington State Transportation Improvement Board, the authors claim: “The Intranet-based dashboard has been the key to restoring financial stability to an agency in dire need of a new approach.” (Gorcester and Reinke, 2007) The authors

state the dashboard supports an active management approach, which has helped avoid problems by analyzing and responding to the warnings the data presents.

## **Standards and Comparisons**

Comparison of performance data with targets is the key step in assessing performance. A long-standing debate about standards seems to be concluding with the notion of benchmarking. A twenty-year-old recommendation of the United Kingdom's Audit Commission of five comparisons for local government gives a good starting point to understand the range of possibilities and the particular advantages of other comparison types (Palmer, 1993):

- Temporal — for example comparing this year's performance with last year's.
- Standards — comparing performance with some standard, which may be derived from local or national statistics or standards.
- Intra-service — comparing the performance of a number of units or sections within a department that provide the same service, for example homes for the elderly.
- Private sector — comparison with provision in the private sector, for example in legal and architectural services.
- Inter-authority — comparisons with all other authorities, all authorities of the same type, or specially selected authorities, which have similar characteristics or with neighboring authorities.

Comparisons to similar units within a department are not always possible, although they do promote competition when appropriate. Comparison of one's own performance on a timeline has been criticized as not capable of "raising the bar," thus not encouraging significant improvement in performance, but it does address self improvement by changing performance compared to prior achievements.

### Managerial Goal Setting

Comparison to a standard or target determined by management is easy to implement and is not fraught with comparability issues. It requires strong management to work well. The management-derived targets can be set with an eye on past performance, or on other jurisdictions, or on where you want to be. The disadvantage is that it is inherently a management perspective and an internal one, at that, because of the lack of firm footing in anything outside the government unit. When the effective manager sets managerially determined goals, he/she encourages improvement but keeps the goals attainable, periodically evaluating the performance measure and the goals. "To be truly effective, standards should be set at a level of performance well above average, but within the bounds of what has been achieved with current best practices and technologies. That is, they should require that agencies and contractors strive for excellence without setting a goal that cannot be achieved." (Richter, 2004)

Several major cities in the United States have developed statistics-based operational “command centers,” based on the police model. One version, CitiStat, is a popular form of managerial goal setting and performance monitoring used throughout the country. Key to a CitiStat approach is comparison, as Behn (2007) states:

It may be the comparison of the city’s data with similar data from other similar municipalities. It may be the comparison of the data for one city unit with the data for another similar unit. It may be the comparison of the data for one city agency with some ideal—an ideal expressed in a general mayoral aspiration or in a specific performance target. Whenever the CitiStat staff reaches any conclusion about the performance of a city agency—positive or negative—they do so based on some kind of comparison.

Also critical to a CitiStat approach is the review process, with many of the governments using CitiStat actually devoting an entire room for the review. They outfit the room to create an air of managerial monitoring and use the room solely for that purpose. The literature reports the Stat models as being very successful, although they are often less analytically sophisticated than other techniques.

Behn continues about how the implementation in Baltimore progressed:

Baltimore launched CitiStat in the summer of 2000. Two years later, Baltimore’s approach to producing results was only beginning to collect real data on results and only beginning to have an impact on its Department of Transportation’s ability to fill potholes quickly. This, however, is not surprising. After all, changing the behavior of any large organization (public, nonprofit, or for-profit) is very difficult. Nevertheless, within two years, Baltimore’s CitiStat was beginning to have the desired impact. Six years later, the impact was significantly bigger.

Queensland saw benefits in performance with CitiStat’s Operational Performance Reviews. A subsequent section of this report on the benefits of using performance measurement provides more details on the Queensland implementation.

The MIDAS system is notable in the literature (Plumridge and Wynnycky, 2007). It is noteworthy for the fact that, like CitiStat, flexibility is exercised in the nature of the comparisons. Ontario implemented the system with a top-down approach, but the users have requested enhanced abilities to be able to compare to other municipalities. The system affords year-to-year comparisons, comparisons to other municipalities for a specific set of results, and comparisons to aggregated performance statistics. Users want the ability to specify the level of comparison and the ability to choose comparable municipalities.

### Inter-jurisdictional Comparisons

Benchmarking, or inter-jurisdictional comparison, is a well-accepted system. In the last twenty years, there has been an increase in the number of governments measuring performance, the number of governments measuring performance using indicators from a central source, and the number of governments mandated to use indicators as a group. All of these factors have made inter-jurisdictional comparisons easier, but they remain less common than comparisons over time. Much of the literature examines the additional benefits available through benchmarking

The increase in the ability to do it practically is not the only reason for the rise in the popularity of benchmarking. It is an inherent desire to compare oneself with others. However, Keehley and MacBride (1997) suggest it offers something no other comparison can. "Any organization can use outcome measures to monitor its internal improvement quarterly, but new horizons can be discovered only through unconventional ideas. This factor is key to benchmarking; without it, breakthrough improvement is impossible." This is echoed with: "By challenging the status quo, the benchmarking process provides public utility managers with a tool that can be used to provide meaningful assessments of an organization's competitive position and point the way to productivity and customer service improvements." (Meszaros and Owen, 1997)

"Local officials would be well advised to face this fact: interjurisdictional comparisons will be made. Those comparisons can be anecdotal, pseudo-systematic (for example, "quick and dirty" studies that often sacrifice precision, consistency, and validity for simplicity and speed), or systematic. The first two types - anecdotal and pseudo-systematic comparisons - rank highest on the cringe-factor scale." (Ammons, 1997)

Unlike managerially set targets, benchmarks are externally determined and free from the power of management. This author finds the automatic nature of the benchmark freeing: "... it serves to create a point of common reference that is based neither on opinions nor on values, but on factual measurements of productivity." (Triantafillou, 2007)

These comparisons are still not without problems, such as those alluded to above in discussing the measures themselves. "The problems with superficial comparisons are numerous. Such comparisons often ignore differences in the nature, scope, and quality of services from place to place." (Ammons et al, 2001) Some of the sources of the problem include the differences in the array of services included in the budget line, cost accounting rules, which costs are included in a budget line, and differences between appropriations and expenditures.

One also needs to account for exogenous factors, which the section on measurement discusses. The reason they are measured is to permit comparisons absent of the effects of these external variables, which government does not control. In addition, governments do set different goals, which specific performance indicators will reflect. The use of profiles captures differences in both environment and goals. New South Wales Department of Local Government (2008) explains how they are used:

When assessing or comparing the performances of councils, it is important to remember that local circumstances can influence how well a council provides its services. There are often good reasons why it is harder or more costly to provide certain services in some local government areas than in others or why a different mix of services may be delivered. In some cases, councils may have made conscious decisions to provide lower or higher levels of services depending on local needs. The council profiles will help you assess the comparative performance information.

In an article on the construction of performance standards, this author also emphasized the concept of standards that are within the manager's control. "...only state and local program efforts-not characteristics of their populations or economic conditions that were beyond program managers' control-should explain why they met, exceeded or fell below their negotiated performance standards." (Courty et al, 2005)

The literature addresses methods for comparing performance between jurisdictions. Performance measurement systems are different from isolated performance measurement initiatives in the fact that they permit comparisons with outside entities. Benchmarking is the act of using outside comparisons. Benchmarking is a management tool to identify better practices (Raaum, 2007). These comparisons with other practices provide a real "base line for performance improvement" (McAdams and O'Neil, 2002).

With comparative data, benchmarking produces a reference that is based on factual level of productivity, rather than historical performance in a single organization. This enables managers to make more meaningful assessments of an organization's performance, using information about its relative performance (Meszaros and Owen, 1997). Triantafillou (2007) reports that benchmarking "urges the organizational performance, which is the target of the comparison, to act." Comparisons "help to establish a performance-based culture in the public sector" (Kouzmin et al, 1999).

Using solely internal historical assessment of performance is not as satisfactory for performance improvement purposes as external benchmarking comparisons (Boyne et al, 2002). However, in the United States, because there are few performance measurement systems, historical comparison is still the main benchmark level found in budget reporting (Willoughby, 2004; Julnes and Holzer, 2008). In a 1999 study of local governments in the U.K., researchers (Boivard and Davis, 1999) concluded, "internal benchmarking activity (sharing ideas around a single authority)<sup>1</sup> appears to be relatively inefficient in many cases." Others (Keehly and MacBride, 1997) went as far as stating that benchmarking has to go further than internal comparisons, otherwise "breakthrough improvement is impossible."

---

<sup>1</sup> The literature uses benchmarking in two other ways in addition to the most common, comparisons against the performance of other jurisdictions. Some authors use it to refer to examining operational differences of other, high-performing organizations and incorporating those changes. It is also used, although infrequently, as any general method of comparison. This is the meaning used by Boivard and Davis in declaring internal benchmarking inefficient.

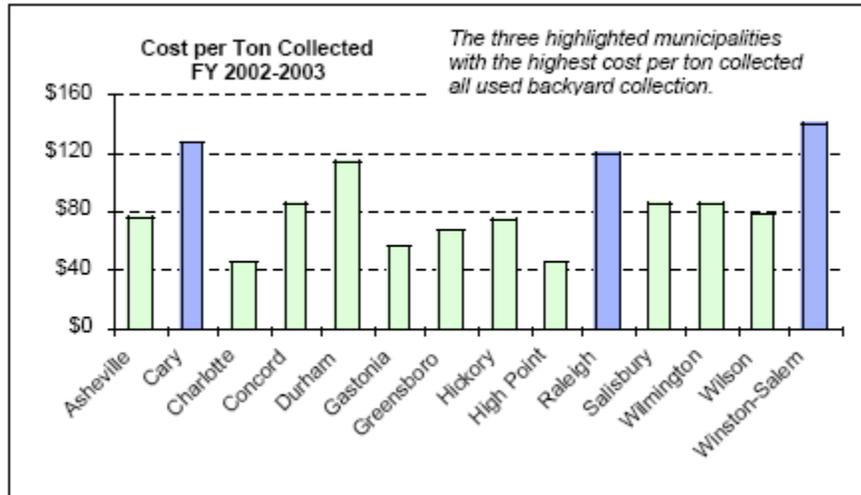
Generally, the literature states that comparisons between similar governments are best. In the empirical results from a nationwide Norwegian benchmarking study, organizational learning is demonstrated as a result of benchmarking (Askim et al, 2008). A key finding of this study conflicts with much of the literature and assumptions of practitioners. Municipalities learn more from dissimilar benchmarking partners. The study found greater learning occurred in networks in which there was greater fiscal heterogeneity among the municipalities.

Comparing performance and costs for different activities with a statewide average is not the best way to identify possible gains in efficiency. Unless professional standards, like police or EMS response time, validate an average, the average is just that (Woodbury et al, 2003) -- it tells little about how good or poor the performance is. Moreover, to identify factors public managers can manipulate to produce better results or services that municipalities might share, one should examine not the typical average case, but the high performing ones to see how they manage differently.

The Best Value approach receives support in some of the literature. A cluster of units comparing their results among themselves seems more effective than a comparison of each unit to the best practice (McAdams and O'Neil, 2002). The study identified mutually supportive programs within the cluster and good guidance material from the government as the causes for the effectiveness of the cluster.

A study of the use of Best Value Performance Plans in Wales finds disappointing results in the improvement in accountability using this approach (Boyne et al, 2002). Pre-existing conditions led to the failure to improve accountability: lack of prior performance indicators and limited staff expertise in performance measurement.

In the North Carolina Project, Winston-Salem was the top performer as a benchmarker, not because it was the best performing municipality in the delivery of services, but because it used the project information to improve its operations to the benefit of citizens. In response to the disappointing result that showed Winston-Salem was most inefficient in providing residential refuse collection service, the city took several steps to reduce its operating cost. As a result, the city was able to reduce the inefficiency factors, and expects to achieve cost savings in upcoming years. (See Figure 1) “True benchmarkers, however, realize that benchmarking is a management tool, not a beauty contest. Benchmarkers select benchmarking partners in hopes of discovering ways to improve their services, not simply as a public relations ploy. They seek out high performance organizations.” (Ammons, 2000)



Estimated Impacts				
Households	Changes in routes	Changes in personnel	Changes in trucks	Estimated savings (\$)
5,000	-1	-3	-1	\$29,000
10,000	-1	-3	-1	-\$14,000
15,000	-2	-8	-2	\$104,000
20,000	-2	-8	-2	\$61,000
25,000	-3	-12	-3	\$144,000

**Figure 1 - Source: North Carolina Local Government Performance Measurement Project. (2005). Benchmarking for Results.**

Another author reviewing the results of the North Carolina Project sees benchmarking as resolving conflicting pressures on public administrators (Williams, 2005). The author perceives the pressure to perform at a high level as being in conflict with the incentives to avoid risks leading to failure and negative publicity. Leveraging performance data to improve program outcomes yields both a high level of performance and avoids risks in performance outcomes.

#### Factors Affecting Benchmarking

There is considerable discussion about how visible benchmarking should be. Early on, it was commonplace to espouse voluntary participation; self-directed analysis; and internal use for management purposes. These and other devices would protect the participant. Now the consensus is that benchmarking is more effective when management makes the results available to the public.

Braadbaart (2007) states publication of performance achievements promotes performance improvement. For example, media coverage and praise from various sides prompted water utilities to develop internal targets, to begin discussing performance at industry seminars, and to swap performance tips. The confidential benchmarking that was taking place between 1989 and 1995 did not even show any improvements over non-

benchmarkers When the benchmarking became public in 1997, the economic performance of the utilities was enhanced and the performance converged.

An article in the Government Finance Review states that government officials were initially apprehensive about reporting some programs had performed below standard (Berman, 2006). Not only did nothing untoward happen to the sitting administrations, but at least one favorable newspaper editorial praised the honesty and forthrightness of the administration. The author suggests “listening” to the public about services to help align government programs to the needs and desires of the citizens who government serves. The recent literature echoes these sentiments about public communication (Gelders et al, 2008; Page and Malinowski, 2004).

Another issue about the use of benchmarking is whether the effectiveness of the tool in promoting program improvement varies along the continuum of voluntary to mandated benchmarking programs. This relates to the discussion of the publication of benchmarking data because of the force behind published materials, but it is also deserving of a separate discussion because a central government can mandate benchmarking, while keeping it as an internal management tool, without external publication.

In Norway, where the political environment is quite different from that in the United States, highly competitive comparisons offer the most improvement in management and performance (Askim et al, 2008). The opposition will use the information on poor performance to wrest power from an opposing party. The use of managed competition to improve performance is not a necessity (Braadbaart, 2007). Collaborative benchmarking can enhance both transparency and performance.

With the Best Value approach to benchmarking, compulsory and defensive modes of benchmarking result from the external accountability that is required (Bowerman et al, 2001). This reduces the focus on tangible improvement. The author’s conclusion is that management of the benchmarking by local government needs to balance the centralizing approach.

Australia found it necessary to mandate performance measurement throughout its local governments (Woodbury et al, 2003). Using the “average municipality” statistic, each state determined what specific service areas should be included for its municipalities. In Ontario, Canada, municipal governments are required to report financial and non-financial information on thirty-five measures of performance in nine core service areas (Chan, 2004).

Proof that voluntary benchmarking does not work is provided by the lack of consistent improvement in the municipalities participating in the North Carolina Project (Williams, 2005). Although there were stars, efficiency gains were not uniformly distributed across all program areas, leaving a great deal of unexplained variance. From a study of health care: “... it is not enough to make use of voluntary benchmarking models to obtain an orientation toward improvement.” (Barretta, 2008)

Successful benchmarkers plan how to use the tool and implement those plans. The less successful incorrectly assume benefits will accrue from having the tool. Hence, in the early stages of the programs for three projects the authors examined in North Carolina, they reported relatively few cost savings or program and service improvements (Ammons et al, 2001).

Best practice efforts are dependent on comprehensive benchmarking and, therefore, may require a mandated system (Bretschneider et al, 2005). The author defines best practice as preferable to any other course of action to achieve a deliberate end. It involves a comparative process, an action, and a linkage between the action and an outcome or goal. If a complete range of comparisons is not available, the notion of a best practice is not defensible, because the comparison was not made to “any other course of action.”

### **Benefits Derived from Performance Measurement**

It may seem frustrating to some, but obvious to others that citizens may not recognize government efforts in performance measurement, benchmarking, and best practices. Furthermore, there is little evidence that service productivity enhances the satisfaction of citizens with their government (Moore et al, 2005). Although Behn (2007) illustrates the cost savings from the control and reduction in the use of overtime due to the CitiStat system in place in Baltimore, he comments: “If you walked through downtown Baltimore and asked individual citizens, “What do you think of CitiStat?” the most honest answer you would get is “Huh?” Citizens don’t pay attention to government’s management strategies. But they care about the results of those strategies.”

CitiStat is a system for improving performance of all departments in a local government, but in the initial implementation in the Police Department of New York City, it was called CompStat. CompStat has had documented benefits in a number of police departments. A recent implementation of CompStat in Queensland provided some detailed documentation of net program benefits (Mazerolle et al, 2007). Queensland documented the cost savings from reduced crime and compared the savings to the costs of over \$1.5 million to conduct the OPR’s (Operational Performance Reviews) since the inception of the program. The overall savings were close to \$1.2 million, after accounting for the costs of running the program. Most of these “costs” were not added costs. Rather, they were the salaries of the participants, which, prior to CompStat, paid for the other management activities they engaged in. The overall savings of \$1.2 million are conservative.

The exciting thing about the Queensland study is the other benefits. Queensland attributed to the OPR’s a reduction of 3,200 crimes of the 11,700 that could have been expected without the introduction of the OPR’s, according to historical trends. This significant decrease in crimes was strongest in the reduction of unlawful entry into properties.

A group of utility providers in the Netherlands reduced costs due to benchmarking (Braadbaart, 2007). Both novice and experienced benchmarkers showed similar pricing behavior, although the novice benchmarkers had the advantage of operating in a mature system.

The results from the North Carolina Project showed a significant reduction in the cost per ton of refuse removal. The officials had been aware of a need to change the recycling and refuse collection systems before they had comparative data, but the hard evidence made the case clearer and provided a needed impetus for change. Data also smoothed the internal negotiations that accompanied the required operational change (Ammons, 2000).

### **A Performance Measurement Model for New Jersey**

The literature review shows a range of different implementations of performance measurement. Academics generally agree that other countries have done more, and there is more sophistication abroad than in the United States. SPAA is not aware of any state that has mandated a performance measurement system, but there are many centralized governments outside of the United States, which have mandated or promoted a centralized system, usually with benchmarking comparisons inherent to it.

One of the most sophisticated and transparent systems is the British Comprehensive Performance Assessment (CPA) system, which has been in place since 2002. Under the CPA systems, British municipalities have to report on a vast number of indicators to the central government, where the data is then fully accessible to citizens. The oldest systems in place are Ontario's Municipal Performance Measures Program (MPMP) and Nova Scotia's Municipal Indicators, which both started in 2000. A third Canadian province, Quebec, started a province-wide municipal performance measurement system in 2003. Different states and territories in Australia and the central government in Norway have mandated municipalities to provide information on certain key service areas. However, there are not as elaborated as the systems presented above.

There are important individual and group implementations in the U.S., including New York City; Worcester, Massachusetts; Portland, Oregon; North Carolina; and Fairfax County, Virginia. More locally, Union Township is proud of its implementation of the CitiStat model.

If New Jersey moves forward to implement the systematic collection of standardized performance measures from all its local governments, it would be the first state in the union to do so. While the effort would be considerable, New Jersey could avoid pitfalls that previous performance measurement system implementations encountered: hastiness of implementation; no or little practitioner outreach in the design phase (Davis, 1998); lack of comparison subcategories (Foltin, 1999) and the absence of shared accounting practices (Coe, 1999).

Most performance measurement systems evolve from their initial implementation. Behn proposes the low hanging fruit approach for a beginning effort: "In this situation, the

mayor might well choose to focus on a few smaller but easily corrected, if not eliminated, problems. Faced with skepticism inside and outside of city government, a mayor could elect to demonstrate some quick wins that can silence the critics and convince others ....” Picking key service areas to begin with and indicators of performance in these areas where improvement is easy to attain would maximize the chance of early success and start a learning process in New Jersey municipalities.

The literature has considered The North Carolina approach to be a good model, but its results in performance improvement are mixed. The key point is that it is a voluntary model that has sought to achieve consensus as it has evolved. Specific suggestions from this project would include using steering committee meetings to discuss new areas of study, building consensus, and ensuring comparability (Rivenbark and Carter, 2000). The state could use this process to evolve a beginning system into a sophisticated system that would improve local government and provide a means to continually measure performance. However, if the incentives and disincentives present in the New Jersey legislation (New Jersey Public Laws Chapter 54, 2007) are not used effectively, it may have the same mixed results as North Carolina experienced.

## COMMENTS ON THE LITERATURE

The data in the literature are thorough and comprehensive. They are a good reflection of the state of our knowledge. The academic community writes extensively on performance measurement. In addition, academics are often the consultants to or creators of systems that are implemented.

The literature has evolved in this field, because it is, as they say, “Hot.” What was written twenty years ago formed a base for the conceptualization of the field and a roadmap to the investigations that followed.

The literature on performance measurement is extensive, but culling out what is most appropriate from the volume is challenging. It is wise to consider the work recently done in other countries when considering what the final model for New Jersey should be. However, New Jersey may need to look elsewhere to find the initial steps it should take to get to a final model.

Bob Behn, who is cited in this review and writes a monthly management series with a performance measurement focus, states that what you need to measure is determined by what you are going to do with it. Determining how the information will be used is an early stage in developing a performance measurement system. Once that decision is known, specific parts of the literature become more appropriate and the designers can formulate a specific model.

No state has centralized a performance measurement system. Countries and sub-units of countries have done so, but the literature only covers what is required to be measured and reported. It is silent on how to manage using such a system.

## CONCLUSIONS

Performance measurement, and the general attempt to measure efficiency of local government, is an increasingly important concern. Both internationally and with increasing frequency in the United States, local governments are implementing performance measurement systems to improve efficiency and outcomes. Overall, the literature points to promising tools or approaches:

- Performance Indicators (Measures)

Performance measures should cover the range of the processes being monitored, from workload and inputs to outputs and outcomes, with a consideration for external variables, which may affect performance potential. Outcomes are most important because they measure those conditions that are concerns of the public and help managers, who accept their responsibilities for outcomes, to think outside the box. One also needs to focus on inputs, internal capacities, and outputs. The literature contains many good measures.

- Data Envelopment Analysis

Analyzing performance data has advanced in recent years. The most powerful tool appears to be Data Envelopment Analysis, which accounts for multiple inputs and outputs and distinguishes between technical, scale, and allocative efficiency. DEA requires sophisticated analytic expertise to use appropriately. However, good management can occur with simple, even basic, forms of data analysis, if it is informed by an understanding of the processes being monitored.

- CitiStat

CitiStat is an example of a successful data review and performance management system, but only one municipality in New Jersey has implemented it. Since it relies more heavily on management review than on sophisticated analysis or benchmarking, it is hard to replicate throughout an entire state. CitiStat does not preclude sophisticated analysis, but it does not rely on it to achieve its internal management goals.

- Benchmarking (Inter-jurisdictional Comparisons)

Local governments should make comparisons in order to maximize improvement in the delivery of services. Benchmarking against the performance of other governments is often a popular choice, but there are difficulties in creating comparable cost-based data. When data is available, stakeholders and managers will benchmark.

- Transparency of Performance Data

Managers and local officials debate whether they should publicize all performance data, but the literature considers the benefits of transparency and critical analysis as overshadowing the concerns of those who fear public reaction

## **APPENDIX A -- ANNOTATED BIBLIOGRAPHY**

### **Literature Review and Analysis Related to Measurement of Local Government Efficiency**

Ammons, D. N. (1997). "Raising the Performance Bar. Locally." Public Management Magazine **79**(9): 10-16.

The article elaborates on the belief that citizens care about government services and discusses how performance measurement can help provide appropriate information to citizens. Efficiency measurement as opposed to resource input measurement can benefit local governments. The potential uses and benefits of efficiency measurement are discussed. The article emphasizes the use of inter-jurisdictional comparisons.

Ammons, D. N. (2000). "Benchmarking as a Performance Management Tool: Experiences Among Municipalities in North Carolina." Journal of Public Budgeting, Accounting & Financial Management **12**(1): 106-124.

The article defines benchmarking. Benchmarking takes 3 distinct forms in the public sector, each serving a different purpose. The first form is corporate benchmarking, which is tied to best practice processes. The second form uses targets as goals. The article investigates the third form of statistical comparisons and concludes that when applied properly and with care, benchmarking is a performance improvement technique that can yield tangible results, as demonstrated by the North Carolina cities of Greensboro, Wilmington, and Winston-Salem.

Ammons, D. N., Charles K. Coe, and Michael Lombardo (2001). "Performance-Comparison Projects in Local Government: Participants' Perspective." Public Administration Review **61**(1): 100-110.

This article examines three prominent projects designed to measure and compare the performance of local governments. All three projects are assessed from the perspective of the local government participants in these projects, revealing gaps between high expectations and subsequent results, but nevertheless suggesting an array of benefits for participants.

Ammons, D. N., and William C. Rivenbark (2008). "Factors Influencing the Use of Performance Data to Improve Municipal Services: Evidence from the North Carolina Benchmarking Project." Public Administration Review **68**(2): 304-318.

Many local governments measure and report their performance, but the record of these governments in actually using performance measures to improve services is more modest. The authors of this study examine patterns of performance measurement use among a set of North Carolina cities and conclude that the types of measures on which officials rely, the willingness of officials to embrace comparison, and the degree to which measures are incorporated into key management systems

distinguish cities that are more likely to use performance measures for service improvement from those less likely to do so.

Askim, J., Åge Johnsen , and Knut-Andreas Christophersen (2008). "Factors Behind Organizational Learning from Benchmarking: Experiences from Norwegian Municipal Benchmarking Networks." Journal of Public Administration Research and Theory **18**(2): 297–320.

Benchmarking rests on the assumption that it supports organizational learning and innovation, but the empirical knowledge that underpins this perceived means-end relationship is limited. This article draws on existing research to develop a framework for analyzing organizational learning outcomes from municipal benchmarking. The framework incorporates explanatory factors at different levels (network and municipality), and with different time perspectives (past and present). The study uses empirical results from a nationwide Norwegian benchmarking project.

Barretta, A. D. (2008). "The Exclusion of Indirect Costs from Efficiency Benchmarking." Benchmarking **15**(4): 345-365.

This study aims to evaluate the effects produced by a strategy aimed at neutralizing one of the “disturbance factors” that may impede the focus on “real (in)efficiency” in relative efficiency assessments within the health-care sector: the exclusion of indirect costs from these comparative analyses. The empirical analysis is based on the statistical elaboration of data from a sample group of hospital sub-units within Italian health care trusts.

Barros, C. P. (2007). "The City and the Police Force:Analysing Relative Efficiency in City Police Precinctswith Data Envelopment Analysis." International Journal of Police Science and Management **9**(2): 164-182.

This paper analyses the efficiency of the Lisbon police force precincts with a 2-stage data envelopment analysis (DEA). In the first stage, the study estimates the DEA efficiency scores and compares the precincts with each other. The aim of this procedure is to seek out those best practices that will lead to the improved performance of all of the precincts. The author ranks the precincts according to their efficiency for the period 2000-2002. In the second stage, he estimates a Tobit model in which the efficiency scores are regressed on socio-economic issues, identifying social causes which vary across the city and that affect deterrence policy. The study considers economic implications of the work.

Behn, R. (2008). "The Performance Target Ethic." Bob Behn's Public Management Report November, 2008; Vol. 6, No. 3 Retrieved February 5, 2009, from <http://www.hks.harvard.edu/thebehnreport/November2008.pdf>.

Part of Bob Behn's monthly public management series, this edition discusses performance targets. The author takes the position that

managerial target setting is not only important, but is a responsibility of elected officials. This position is different than, but not totally incompatible with benchmark targets based on inter-jurisdictional comparisons.

Behn, R. D. (2007). What All Mayors Would Like to Know About Baltimore's CitiStat Performance Strategy. Managing for Performance and Results Series. Washington, DC, IBM Center for The Business of Government: 1 - 59.

This report summarizes and presents the questions most frequently posed to CitiStat staff and to Mayor Martin O'Malley by visiting mayors who want to investigate the success of CitiStat. When CitiStat is viewed as a leadership strategy rather than a management system, Dr. Behn argues, the program can be replicated and customized to each mayor's individual needs and priorities. A key insight from this report is that there is no single, right approach as to how to develop a successful management performance and accountability structure. Success depends heavily on clear goals, committed leadership, and persistent follow-up.

Berman, B. J. C. (2006). "The Voices of the People: Missing Links in Performance Measurement and Management." Government Finance Review **22**(3): 16-20.

The purpose of this article is to convey to readers the importance of listening to and understanding how the public judges governmental performance so that you can develop, review, produce, and report performance measures. Failing to involve the public in performance measurement and reporting can lead to dissonance between the government and its constituents, and result in misalignment of government programs.

Boivard, T., and Paul Davis (1999). "Learning to Manage within Limited Resources Coping Strategies and Learning Breakthroughs in UK Local Government." The International Journal of Public Sector Management **12**(3): 293-313.

This article outlines the main conclusions of a research project into how United Kingdom local authorities are managing within limited resources (MLR). Frameworks are developed to aid authorities to plan their approaches to MLR and to situate what they have already done and what they plan to do within a wider portfolio of tactics and strategies. An evaluation is made of how well local government is learning its way through to getting "more from less" and of what local authority support agencies need to do to help authorities to accelerate their learning. Finally, the authors argue that existing learning systems like benchmarking and quality management, while developing rapidly in local government, need further significant refinement if the costs and benefits of resource management strategies are to be systematically evaluated.

Bowerman, M., Amanda Ball, and Graham Francis (2001). "Benchmarking as a Tool for the Modernisation of Local Government." Financial Accountability and Management **17**(4): 321-329.

This paper examines benchmarking as a tool of the modernization process in local government. The contradictory tensions in the Best Value scheme are explored. Benchmarking is shown to encapsulate the dichotomous nature of a modernizing philosophy which espouses innovation and local solutions alongside the government's centralizing tendencies. One consequence is the advancement of 'compulsory' and 'defensive' modes of benchmarking with local authorities benchmarking for external accountability reasons; issues of tangible improvement are secondary concerns. These tensions are viewed as irreconcilable, the implication is that local government will need to carefully manage and evaluate its benchmarking activities.

Boyne, G., Julian Gould-Williams, Jennifer Law, and Richard Walker (2002). "Plans, Performance Information and Accountability: The Case of Best Value." Public Administration **80**(4): 691-710.

The current UK government emphasizes the importance of mechanisms of accountability that involve the planning and public reporting of performance. One example of this is the Best Value performance plan. However, there has been little evaluation of the quality of the information provided in this type of document. This paper draws on literature on stake holding and user needs to identify the data required for accountability. It then assesses whether the plans produced by Best Value pilot authorities in Wales provide appropriate information. The analysis shows that very few of the plans contained the relevant material. Interviews in the pilot authorities highlighted two key reasons for the poor level of data: a lack of performance indicators prior to Best Value and limited staff expertise in performance measurement. The evidence suggests that documents such as performance plans currently make little contribution to the accountability of public organizations.

Boyne, G. A. (1996). "Scale, Performance and the New Public Management: An Empirical Analysis of Local Authority Services." Journal of Management Studies **33**(6): 809-826.

New public management (NPM) arguments on strategy and structure suggest that performance is enhanced if large organizations are disaggregated into smaller units. The NPM perspective reflects the views of public choice theorists who claim that big organizations are unresponsive to public needs, inefficient and fail to achieve their formal goals. These arguments have underpinned many recent changes in the structure of public services at both central and local levels. This paper uses data on six local government services to test the NPM hypothesis that there is a negative relationship between scale and performance. Five dimensions of performance are analyzed: service coverage, quality, speed

of provision, efficiency, and administrative effectiveness. Scale is measured through indicators of service output, caseload and needs. The impact of scale is tested in multivariate statistical models which control for other potential influences on variations in performance across local authorities.

Braadbaart, O. (2007). "Collaborative Benchmarking, Transparency and Performance: Evidence from The Netherlands Water Supply Industry." Benchmarking: An International Journal **14**(6): 677-692.

The purpose of this paper is to investigate how benchmarking affects transparency and economic performance in the public sector. The paper applies a quasi-experimental method to 1989-2000 time series data on benchmarking and non-benchmarking water utilities in the Netherlands. The main limitation of this study is its single-industry scope.

Bretschneider, S., Frederick J. Marc-Aurele, Jiannan Wu (2005). ""Best Practices" Research: A Methodological Guide for the Perplexed." Journal of Public Administration Research and Theory **15**(2): 307-323.

The purpose of this article is two-fold. First, it critically examines the underlying assumptions associated with "best practices research" in Public Administration in order to distill an appropriate set of rules to frame research designs for best practice studies. Second, it reviews several statistical approaches that provide a rigorous empirical basis for identification of "best practices" in public organizations - methods for modeling extreme behavior (i.e., iteratively weighted least squares and quartile regression) and measuring relative technical efficiency (data envelopment analysis, known as DEA).

Chalos, P., and Joseph Cherian (1995). "An Application of Data Envelopment Analysis to Public Sector Performance Measurement and Accountability." Journal of Accounting and Public Policy **14**(2): 143-160.

Public sector performance measures have been advocated by numerous regulatory bodies (FASB 1980; CCAF 1987; CICA 1988; GASB 1989). Proposed Service Effort and Accomplishment (SEA) disclosures include measures of inputs and input-output efficiency ratios (GASB 1989). Such metrics, it is argued, improve internal control and external accountability. To date, SEA disclosures are not widely disseminated. The study examined the potential utility of SEA disclosures in determining public policy in education. Specific concerns of the Illinois State Board of Education and the state legislature with respect to fiscal neutrality and tax referenda in education were examined. As proposed by the GASB (1989, x-xii), educational input and output indicators were compiled for 207 school districts. To overcome specification problems, DEA analysis was used to measure input-output efficiency. Non-parametric hypothesis tests were performed on the resulting efficiency and effectiveness measures.

Chan, Y.-C. L. (2004). "Performance Measurement and Adoption of Balanced Scorecards: A Survey of Municipal Governments in the USA and Canada." International Journal of Public Sector Management **17**(3): 204-221.

To deal with financial constraints and increasing demand on accountability, government administrators have begun implementing modern management tools in their organizations. The balanced scorecard, a performance and strategic management system, has been adopted in for-profit organizations with success and its application in the government sector is explored in this study. Results of a survey of municipal governments in the USA and Canada show that there is limited use of the balanced scorecard. Most municipal governments, however, have developed measures to assess their organizations' financial, customer satisfaction, operating efficiency, innovation and change, and employee performance.

Cheung, R. (2008). "The Interaction between Public and Private Governments: An Empirical Analysis." Journal of Urban Economics **63** (3): 885-901.

Private governments, found in planned developments and condominiums, are increasingly common methods of delivering local services to residents. This paper provides the first empirical study of their impact on local public finance. A novel data set of homeowners' associations allows construction of a panel of private governments in California. Panel methods test whether public expenditures respond to private government prevalence. The study employs a very large data set, 110 cities with 38-year time series, yielding reasonably convincing results.

Claps, M. (2008a). Case Study: Business Process Analysis Underlies Launch of Local Government Shared Service in the U.K.: 1-8.

This article is a case study set in U.K. regarding shared services and partnerships when three local U.K. councils came together to reduce cost and share services. Taxes and revenue sharing are discussed as well as the intricacies of U.K. partnerships. Different stages of the partnering are discussed as well as the success factors and lessons learned.

Coe, C. (1999). "Local Government Benchmarking: Lessons from Two Major Multigovernment Efforts." Public Administration Review **59**(2): 110-123.

Benchmarking and performance measurement are increasingly "hot" subjects among public administrators. Among research based efforts encouraged by the performance measurement drive are two projects that attempt to develop uniform measures so that managers can compare the performance of different city and county governments. This article compares the methodology used in the ICMA and IOG projects. It begins with a brief overview of each benchmarking effort and then looks at the often unexpected obstacles the projects encountered and how they attempted to surmount them. It concludes with some broader lessons that these experiences suggest for local government benchmarking.

Courty, P., Carolyn Heinrich, and Gerald Marschke (2005). "Setting the Standard in Performance Measurement Systems." International Public Management Journal **8**(3): 321-347.

A fundamental challenge in the design of performance measurement and incentive systems is the establishment of appropriate benchmark levels of performance, also known as performance standards. Drawing from the information economics, contract theory and public administration literatures, the authors derive theoretical implications for the construction of performance standards. They then assess alternative methods that are commonly used to construct performance standards and consider their application in performance measurement systems in public programs. The authors draw out important lessons for the establishment of performance benchmarks and other implications for performance standards system design in public organizations.

Davis, P. (1998). "The Burgeoning of Benchmarking in British Local Government: The Value of "Learning by Looking" in the Public Services." Benchmarking for Quality Management & Technology **5**(4).

A large number of benchmarking initiatives have now been established in UK local government and, in most cases, these activities have grown organically from a variety of local pilot schemes and from the social interaction fostered through professional networks, to the point where some classification and clarification is possible. This growth has been powered by a small number of key stimuli. The "Inter Authorities Group" (IAG), the largest voluntary benchmarking club in local government, emerged from the concern over consistent costing data. This article evaluates the current status of benchmarking in UK local government.

Drake, L., and R. Simper (2000). "Productivity Estimation and the Size-Efficiency Relationship in English and Welsh Police Forces: An Application of DEA and Multiple Discriminant Analysis." International Review of Law and Economics **20**(1): 53-73.

This article utilizes data envelopment analysis (DEA) to estimate the productivity of the English and Welsh police forces and to determine whether there are categorical scale effects in policing using multiple discriminant analysis (MDA). The article demonstrates that by using DEA efficiency results it is possible to make inferences about the optimal size and structure of the English and Welsh police forces.

Drake, L., and R. Simper (2002). "X-Efficiency and Scale Economies in Policing: A Comparative Study Using the Distribution Free Approach and DEA." Applied Economics **34**(15): 1859-1870.

This study uses both parametric and nonparametric techniques to analyze scale economies and relative efficiency levels in policing in England and Wales. Both techniques suggest the presence of significant scale effects

in policing and considerable divergence in relative efficiency levels across police forces.

Eagle, K. (2004). "Translating Strategy: Public Sector Applications of the Balanced Scorecard." Government Finance Review: 16-22.

This article discusses the evolution of strategy development and performance measurement in the City of Charlotte, with an emphasis on the lessons learned from nearly 10 years of using the Balanced Scorecard model. Developed for the private sector in an effort to balance measures of organizational performance between financial results and other critical success factors, the Balanced Scorecard has been adapted for the public sector by a number of local governments. From implementing and using the Balanced Scorecard to the most recent challenge of linking strategy to resource allocation, Charlotte's story is one characterized by continuous improvement.

Foltin, C. (1999). "State and Local Government Performance: It's Time to Measure Up!" The Government Accountants Journal **48**(1): 40-46.

State and local governments must do more with less. This has been the task for nearly three decades. Resources have decreased due to slowed population growth, recessions in the 1970s and 1980s, reduced federal funding and reluctance of citizens to pass new taxes. Furthermore, unfunded mandates, inflation, worn out infrastructure and increasing wage and benefit costs have caused expenditures to skyrocket. In light of these conditions this paper encourages the use of performance measurement and provides help in understanding the basic concepts.

Fry, J. C. (2004). Performance Measurement - The Basics. New Jersey League Of Municipalities Annual Conference. Atlantic City, NJ.

The PowerPoint presentation was sponsored by the New Jersey League of Municipalities and the New Jersey Municipal Managers Association. The presentation was oriented to an audience of practitioners in municipal government who had limited experience with performance measurement. It defines performance measurement, provides the basic steps to do it, describes the benefits, lists the types of indicators, provides sources of indicators, elaborates on types of targets, and discusses what should be done to achieve results.

Garcia-Sanchez, I. (2006). "Efficiency Measurement in Spanish Local Government: The Case of Municipal Water Services." Review of Policy Research **23**(2): 355-371.

In this study, the authors established a procedure for evaluating the efficiency of providing the water supply. This procedure has allowed the authors to find that the proposed indicators have a discriminating capability in the analysis of the service, and to reject criticisms traditionally assigned to the sensitivity of the data envelopment analysis (DEA)

technique in relation to degrees of freedom. The article studies efficiency but is also illustrative of the use of the technique of Data Envelopment Analysis.

Gelders, D., Mirjam Galetzka, Jan Pieter Verckens, and Erwin Seydel (2008). "Showing Results? An Analysis of the Perceptions of Internal and External Stakeholders of the Public Performance Communication by the Belgian and Dutch Railways." Government Information Quarterly **25**(2): 221–238.

Performance measurement and communicating about it with the broader public is not self-evident if one looks at public services organizations. In Belgium and the Netherlands, one organization that has been under constant surveillance from its stakeholders is the railroad company. Originally a national public service, it has changed through a European directive to operate in a liberalized transportation market. In this paper, the authors present the results of a qualitative study into what the Belgian and Dutch railway companies measure about their performance towards travelers, how they measure it, and specifically how they communicate it to their employees and stakeholders. Although we might expect the Dutch railway company (NS) to be more transparent than the Belgian railway company (NMBS/SNCB), in reality the two organizations do not differ that much from each other.

Gorcester, S., and Rhonda Reinke (2007). "Dashboards Simplify Performance Reporting." Government Finance Review **23**(5): 53-56.

Successful performance management demands effective performance reporting. Software dashboards, in particular, offer elegant simplicity to the otherwise difficult task of reporting performance for quick interpretation by both the numbers people and the ideas people. The Washington State Transportation Improvement Board (TIB) implemented its performance dashboard in 2003 and followed the government management, accountability and performance directive (GMAP) in earnest after its adoption in 2004 by the State of Washington.

Government Accounting Standards Board (1994). Concepts Statement No.2, Service Efforts and Accomplishments Reporting

In 1994, after extensive research on SEA, the GASB issued Concepts Statement No. 2, Service Efforts and Accomplishments Reporting. That Statement provided the opportunity to elaborate on accountability in government and the importance of reporting SEA performance information as part of general purpose external financial reporting (GPEFR). The Board stated that SEA performance information is necessary for assessing accountability and in making informed decisions: "Therefore, to be more complete, general purpose external financial reporting needs to include service efforts and accomplishments information." This report is not included in the LUARCC articles folder, although it may be purchased from GASB. GASB has continued from this point, particularly

emphasizing citizen involvement in determining what should be measured and reported upon. In 2008, the Board issued an Exposure Draft of an amendment of Concepts Statement No. 2.

Hatry, H., and Joseph J. Wholey (1999). Performance Measurement: Getting Results, Urban Institute Press.

Long before "reinventing government" came into vogue, the Urban Institute pioneered methods for government and human services agencies to measure the performance of their programs. This book synthesizes more than two decades of Harry Hatry's pioneering work on performance measurement into a comprehensive guide. The author explains every component of the process, from identifying the program's mission, objectives, customers, and trackable outcomes to finding the best indicators for each outcome, the sources of data, and how to collect them. He covers the selection of indicator breakouts and benchmarks for comparisons to actual values and suggests a number of uses for performance information. Joseph Wholey contributes a chapter on maintaining the quality of the performance measurement process. It is not available in the Articles Folder for the LUARCC project.

Ho, S.-J. K., and Yee-Ching Lilian Chan (2002). "Performance Measurement and the Implementation of Balanced Scorecards in Municipal Governments." Journal of Government Financial Management 51(4): 8-19.

Since the Government Accounting Standards Board released Concepts Statement No. 2, Service Efforts and Accomplishments (SEA) Reporting in 1994, about one-third of local governments participate in SEA reporting, according to a recent survey. Government administrators have applied some modern, private sector management tools to solve their management problems. However, the Balanced Scorecard, a performance management system used by about 50 percent of Fortune 1000 companies, according to a recent survey by Bain & Co., has not been well received by local governments. As a follow-up to previous studies, including one by Ho and Kidwell that showed the Balanced Scorecard to be the least prevalent tool, this article examines the current performance measurement systems of municipal governments, assesses their perceived value and quality, and compares the experience of a Balanced Scorecard implementer with that of other local municipalities. This article is not available in the articles folder for the LUARCC project but is available for purchase from the AGA.

International City/County Management Association (2008). What Works: How Local Governments Have Made the Leap from Measurement to Management, ICMA Press.

This book was written by the staff of the Center for Performance Measurement at ICMA (International City/County Management Association). The book reports how performance measurement promotes

positive change using over 70 case studies in fifteen different services areas, all from local government. The Center for Performance Measurements runs the comparative performance measurement data collection effort that is now over fifteen years old and has many local governments throughout the United States as its consortium members. The book is not in the articles folder for the LUARCC project, but is available for purchase from ICMA.

Julnes, P. d. L., and Marc Holzer, Ed. (2008). Performance Measurement: Building Theory, Improving Practice, M.E. Sharpe, Inc.

This volume in the ASPA Classics series compiles the most influential contributions to the theory and practice of performance measurement that have been published in various journals affiliated with the American Society for Public Administration. The book includes major sections of original text along with the readings, and provides students and practitioners with a handy reference source for theory development and practice improvement in performance measurement. The coverage is broad, including methods and techniques for developing effective performance measurement systems, building performance-based management systems, and sustaining performance-based budgeting. The articles are all classics in the field that have endured the test of time and are considered "must reads" on performance measurement. This book is not available in the Articles Folder for the LUARCC project.

Keehley, P., and Sue A. MacBride (1997). "Can Benchmarking for Best Practices work for Government?" Quality Progress **30**(3): 75-80.

This article makes the case for benchmarking in the public sector. The improvements in quality and efficiency obtained by private industry through best practices and benchmarking have been so profound and widespread that public sector agencies - organizations traditionally perceived as slow to plan and implement change - are waking up to the benefits of this powerful analytical tool. The article reviews the New York City Transit Authority's adoption of best practices. It also provides ten iterative steps to take in order to benchmark.

Kouzmin, A., Elke Loffler, Helmut Klages, and Nada Korac-Kakabadse (1999). "Benchmarking and Performance Measurement in Public Sectors Towards Learning for Agency Effectiveness." International Journal of Public Sector Management **12**(2).

There are some encouraging examples of benchmarking within the public sector. This paper critically analyzes these examples in order to establish the vulnerability points of such measurement instruments. Additional research is in order to establish the specific learning dimensions of benchmarking and to illustrate the importance of such benchmarking and learning within the highly risky, information technology (IT)-driven experiences of systems development and failure. The paper also

compares the public sector motivation for benchmarking to that in the private sector, where it is better established, to provide insight into differences that may be important in public sector use of benchmarking.

Mante, B., and Greg O'Brien (2002). "Efficiency Measurement of Australian Public Sector Organisations: The Case of State Secondary Schools in Victoria." Journal of Educational Administration **40**(3): 274-296.

This paper provides a review and an illustration of the Data Envelopment Analysis (DEA) methodology for measuring the relative efficiency of public sector organizations performing similar tasks. The paper focuses on measuring the relative technical efficiency of State secondary schools in a geographical region in the Australian State of Victoria. It recognizes that state secondary schools, like other non-profit making organizations, produce multiple outcomes by combining alternative discretionary and non-discretionary inputs.

Mazerolle, L., Sacha Rombouts, and James McBroom (2007). "The Impact of COMPSTAT on Reported Crime in Queensland." International Journal of Police Strategies & Management **30**(2): 237-256.

The purpose of this paper is to evaluate the impact of Queensland (Australia) Police Service's version of COMPSTAT, known as "Operational Performance Reviews" (OPRs), on reported crime. The paper employed interrupted time series analytic techniques to examine the impact of OPRs on various categories of reported crime in Queensland. The analyses assessed the extent to which OPRs were associated with crime reductions across the 29 police districts in Queensland. This study was the first to provide a systematic examination of the impact of OPRs on a relatively large number of offence categories.

McAdams, R., and Liam O'Neil (2002). "Evaluating Best Value through Clustered Benchmarking in UK Local Government: Building Control Services." International Journal of Public Sector Management **15**(6): 438-457.

The Best Value initiative has been designated as the key framework for improving service quality and effectiveness in UK local government. In support of the framework a number of measures have been introduced to measure effectiveness. However, there are problems with using comparative performance measures within the diverse grouping of local government services. This article measures the effectiveness of Best Value in a similar group of building control services by using a clustered benchmarking approach. The mainly qualitative research involves applying a strategic benchmarking approach to the 26 units or councils within the cluster and then comparing the cluster with the best practice UK measures for building control services. To facilitate the process a benchmarking measurement framework for building control was developed.

Meszaros, G., and C. James Owen (1997). "Improving through Benchmarking: A Competitive Assessment of Municipal Utility Services." Public Works Management & Policy 2(1): 11-23.

This article examines the use of competitive benchmarking as a means of assessing and responding to the changing market conditions sweeping the municipal water and wastewater utility industry. The authors present an overview of the benchmarking process, and illustrate the use of benchmarking through an examination of the techniques used to develop a competitive assessment of the Utility Division of the city of Fort Wayne. The authors perceive that rising public expectations for higher-quality, more cost-effective services are radically altering the way municipalities deliver utility services. In increasing numbers, many local governments are introducing competition, through private sector contract operation companies, into their water and wastewater utilities. Once foreign to public utility managers, the concepts, behavior, and work practices of a competitive market increasingly are becoming a reality for many municipal water and wastewater utility providers.

Moore, A., James Nolan, and Geoffrey F. Segal (2005). "Putting Out the Trash: Measuring Municipal Service Efficiency in U.S. Cities." Urban Affairs Review 41(2): 237-259.

The authors discuss Data envelopment analysis (DEA) as a response to their view that the literature describing the performance of municipal services often uses imperfect or partial measures of efficiency. DEA has emerged as an effective tool for measuring the relative efficiency of public service provision. This article uses DEA to measure the relative efficiency of 11 municipal services in 46 of the largest cities in the United States over a period of 6 years. In addition, this information is used to explore efficiency differences between cities and services and provide input into a statistical analysis to explore factors that may explain differences in efficiency between cities. Finally, the authors discuss municipal governments' use of performance measures and problems with collecting municipal data for benchmarking.

Morton, L. W., Yu-Che Chen, and Ricardo S. Morse (2008). "Small Town Civic Structure and Interlocal Collaboration for Public Services." City and Community 7(1): 45-60.

Local governments are responsible for financing and providing an array of public services ranging from police, fire, and emergency medical services to streets, parks, and water. Two mechanisms, namely civic structure and interlocal collaboration, have the potential to solve the problem of providing high quality public services in the face of declining resources and increasing needs. The authors find that civic structure - citizen engagement in solving public problems - is positively and strongly associated with perceived quality of small town public services. Although many rural towns show significant positive association with citizen ratings

of overall service quality, citizens seem to prefer their local government directly providing police services rather than entering into interlocal agreements. This suggests that many services are better provided directly.

New Jersey Public Laws (2007). P.L. 2007, Chapter 54. NJSA 52:27D-181.1. N. J. S. Government.

Chapter 54 is an act establishing the "Local Unit Alignment, Reorganization, and Consolidation Commission," performance measures for municipalities, and the Municipal Efficiency Promotion Aid Program compiled at N.J.S.A. 52:27D et seq. It also pertains to the development of performance measures by the New Jersey State Local Finance Board. Section 181 addresses how the performance measures shall be used to distribute aid to municipalities that meet the performance measures promulgated by the Board. The articles folder for the LUARCC project contains the amended statute found in section 181.1. For reference, the file also includes N.J.S.A. 52:27D-18.2. , P.L. 2007, Chapter 54 and an excerpt from the original bill A-50, introduced by Senator Roberts in the 2006 legislative session.

New South Wales Department of Local Government (2008). Comparative Information on NSW Local Government Councils 2005/2006. Sydney, NSW, Australia, New South Wales Department of Local Government: 245p.

This publication of the New South Wales (Australia) Department of Local Government provides detailed performance information for three years for every municipality in New South Wales. This publication uses municipal profiles to group councils into similar groups for comparison. It represents an example of public benchmarking.

Ouellette, P., and Valérie Vierstraete (2005). "An Evaluation of the Efficiency of Québec's School Boards Using the Data Envelopment Analysis Method." Applied Economics **37**(14): 1643-1653.

In this paper the efficiency of Quebec's school boards during a period of severe cutbacks in their finance is examined using Data Envelopment Analysis. The average efficiency is found to be relatively high. In spite of this, potential savings could be achieved if school boards were fully efficient. Results depended heavily on school boards' socio-economic conditions. They were subjected to Tobit analysis and the boards' corrected efficiencies recalculated. The inefficiencies cost \$800 million of which \$200 million came from unfavorable socio-economic conditions.

Pachon, H. P., and Nicholas P. Lovrich, Jr. (1977). "The Consolidation of Urban Public Services: A Focus on the Police." Public Administration Review **37**(6): 38-47.

A set of literature in public administration has recently emerged, associated with the Public Choice perspective, which argues against

metropolitan consolidation and police consolidation specifically. This article reviews the major assertions which the "anti-consolidationists" make in specific reference to the consolidation of urban police departments: (1) that larger police departments are not more efficient or economical; and (2) that citizen satisfaction with urban police services varies inversely with the size of the municipality. On the first assertion the authors find that evidence is still inconclusive regarding the economy and efficiency of larger police departments. On the second assertion, the results indicate that, contrary to the anti-consolidationist viewpoint, citizen satisfaction with police services is more a reflection of the differentiated socio-economic characteristics of central cities and suburban cities than a function of city size.

Page, S., and Chris Malinowski (2004). "Top 10 Performance Measurement Do's and Don'ts." Government Finance Review **20**(5): 28-32.

Performance measurement continues to gain in popularity in all levels and types of governmental organizations. Robert Kaplan's Balanced Scorecard, David Osborne's Reinventing Government, the success of the NYPD's COMPSTAT process, and the ever-increasing public pressure to report performance have made the practice of performance measurement commonplace in the public sector. While most of these efforts are well intentioned, too many performance measurement programs fail to live up to expectations or get derailed altogether because of poor execution. To help governments make the most of their performance measurement efforts, the authors compiled a "top 10 list" of performance measurement do's and don'ts, using case studies to illustrate their assertions.

Palmer, A. J. (1993). "Performance Measurement in Local Government." Public Money & Management **13**(4): 31-36.

Emphasis on improving performance in the public sector has led to the development of systems for measuring performance in local government. This article reports on the author's research into this development, presenting information on how local authorities measure performance; which dimensions of performance they attempt to measure; and how they use performance measurement information. Existing systems of measuring performance are evaluated in relation to an ideal system. The conclusion is that, although few authorities currently have a coherent system for measuring performance, pressure for continued development of such systems should yield improvements.

Pedraja-Chaparro, F., and Javier Salinas-Jiminez (1996). "An Assessment of the Efficiency of Spanish Courts Using DEA." Applied Economics **28**: 1391-1403.

The objective of the article is to provide a measure of technical efficiency of the Administrative Litigation Division of the Spanish High Courts. The concept of efficiency to be measured and the most adequate technique for carrying out the efficiency analysis are selected by considering the specific

characteristics of public production. The analysis is undertaken by using data envelopment analysis (DEA) and various homogeneity tests (returns to scale and restrictions on weights) are applied in order to ensure a correct comparison between Courts.

Plumridge, N., and Bohdan Wynnycky (2007). "The "Midas" Touch: New System Will Bring Powerful Analysis Tools Online for Ontario Municipalities." Municipal World **117**(10): 27-30.

As a response to the growing desire to access more financial and performance measurement information, the Association of Municipalities of Ontario (AMO) and the Ministry of Municipal Affairs and Housing (MMAH), with funding assistance from the Ministry of Finance's (MOF) "Strengthening Our Partnerships" program, jointly developed the Municipal Information and Data Analysis System (MIDAS). MIDAS is a powerful web-based query and analysis tool that will allow any municipal staff member to access the data compiled in the Province of Ontario's Financial Information Return (FIR) system, including the Municipal Performance Measurement Program (MPMP) data. Staff members and elected officials can query and report on this statistical, financial and operational data, examine their own historical records, and browse and make comparisons with their peer municipalities using standard and established performance measures and indicators. This increased access to financial and performance measurement data allows for an improved ability to make service delivery decisions.

Public Performance Measurement and Reporting Network (2009). "Public Performance Measurement and Reporting Network." Retrieved February 3, 2009, from [www.PPMRN.net](http://www.PPMRN.net).

This website serves as a resource for the Rutgers University School of Public Affairs and Administration's network of over 2,000 academics and practitioners. It is available to the public and contains information and other information resources on the field of performance measurement and related topics. The goal of the Public Performance Measurement and Reporting Network is to promote the use of valid, reliable data as a key element in improving the delivery of public services. In support of the Network, the National Center for Public Performance has implemented a series of initiatives: a comprehensive and continuously updated database of publications and cases; national conferences and workshops; publications of measurement-based books and articles; an Online Public Performance Measurement Certificate; and a monthly e-newsletter.

Raum, R. B. (2007). "Applying the Finding Paradigm to Audits of Output and Outcome Performance: A Research Perspective." Journal of Government Financial Management **56**(2): 44-50.

The "finding paradigm," a four-element measurement tool used to analyze government service provision, is at the foundation of one method of

performance auditing. The four-element finding paradigm of criteria, condition, effect and cause is widely used by auditors and cited in auditing standards, and has proven a powerful tool for analyzing the performance of all manner of government, as well as private and nonprofit operations and programs. This article provides guidance and details for applying the finding paradigm to findings for audits of output and outcome performance.

Revelli, F., and Per Tovmo (2007). "Revealed Yardstick Competition: Local Government Efficiency Patterns in Norway." Journal of Urban Economics **62**(1): 121–134.

This paper investigates whether the production efficiency of Norwegian local governments exhibits a spatial pattern that is compatible with the hypothesis of yardstick competition. In order to check whether yardstick competition is really responsible for the observed spatial pattern and rule out alternative theoretical explanations, the paper exploits unique information from a survey on local politicians' attitudes towards comparative evaluation of local bureaus' performances against other jurisdictions' (benchmarking).

Richter, C. A. (2004). "The Case for Performance Standards." Public Roads **67**(5): 18-22.

This article makes the case for performance standards, expressed from a user perspective, as opposed to specifications in the construction of highways. The author suggested customer focused issues include street smoothness, road noise, longevity of the road, and traffic congestion. The article reviews the use of performance standards in practice throughout the United States.

Ridley, C. E. (1927). Measuring Municipal Government, School of Citizenship and Public Affairs, Syracuse University.

This seminal work in the field of performance measurement is referenced for the LUARCC projects to provide a sense of the long history of attempting to measure the efficiency of government. The book is not provided in the Article Folder for the LUARCC project and is out of print but the concepts are the basis of much of what is being implemented and studied today.

Rivenbark, W. C., and K. Lee Carter (2000). "Benchmarking and Cost Accounting: The North Carolina Approach." Journal of Public Budgeting, Accounting & Financial Management **12**(1): 125-137.

Benchmarking is a management tool that promotes process improvement. By comparing services units across jurisdictions, best practices can be identified and used to enhance less efficient and effective operations. However, the lack of generally accepted criteria to compare service costs for local government has hindered benchmarking initiatives. One of the

key components of the North Carolina Local Government Performance Measurement Project is the full-cost accounting model developed to ensure that localities employ the same methodology to collect and report cost data associated with performance measures. This article presents an overview of the development and implementation issues associated with that model and highlights the areas of direct costs, indirect costs, and capital costs. It is argued that accuracy and comparability of performance and cost data are the fundamental ingredients of a benchmarking and performance measurement project.

Schochet, P. Z., and J. A. Burghardt (2008). "Do Job Corps Performance Measures Track Program Impacts?" Journal of Policy Analysis and Management **27** (3): 556-576.

Since the 1993 Government Performance and Results Act, performance measurement systems based on short-term program outcomes have been increasingly used to assess the effectiveness of federal programs. This paper examines the association between program performance measures and long-term program impacts, using nine-year follow-up data from a recent large scale, national experimental evaluation of Job Corps, the nation's largest federal job training program for disadvantaged youths. Job Corps is an important test case because it uses a comprehensive performance system that is widely emulated. Randomized experimental design was employed. That is, the assessment of program impacts is compared between those who received services and those who did not. This article focuses on a federal agency.

Stevens, P. A. (2005). "Assessing the Performance of Local Government." National Institute Economic Review **193**(1): 90-101.

This article considers the measurement of performance in the public sector in general, focusing on local government and the provision of library services by English local authorities in particular. The author considers two methodologies that assess the performance of local authorities in terms of the efficiency with which they provide services, and considers methods that allow for the identification of exogenous influences on performance such as the socio-economic profile of the population served by the authority. The author finds that although both methods' results appear similar, the implications for potential cost savings vary widely. Omitting to account for background factors leads to an overstatement of the level of inefficiency and hence the scope for reducing expenditure.

Sun, S. (2002). "Measuring the Relative Efficiency of Police Precincts Using Data Envelopment Analysis." Socio-Economic Planning Sciences **36**(1): 51-71.

Data envelopment analysis (DEA) is used to measure the relative efficiency of the 14 police precincts in Taipei city, Taiwan. The results indicate how DEA may be used to evaluate these police precincts from commonly available police statistical data for the years 1994-1996. To

sharpen the efficiency estimates, the study uses window analysis, slack variable analysis, and output-oriented DEA models with both constant and variable returns to scale. The problem of the presence of non-discretionary input variables is explicitly treated in the models used. Potential improvements in technical efficiency of police precincts are examined by readjusting the particular output/input indicators. The analysis indicates that differences in operating environments, such as resident population and location factors, do not have a significant influence upon the efficiency of police precincts.

Triantafyllou, P. (2007). "Benchmarking in the Public Sector: A Critical Conceptual Framework." Public Administration **85**(3): 829–846.

How can we critically address benchmarking? By conceptualizing benchmarking as a normalizing governing technology, a space is cleared for analyzing some of the power relations brought into play in benchmarking activities. As a device of power, benchmarking depends upon the production of normalizing knowledge and the freedom or self-governing capacities of those who are benchmarked. The fruitfulness of this conception is illustrated through an example from the Danish hospital system.

United States Conference of Mayors (2000). "Project Scorecard Evaluates Cleanliness of New York Streets." Retrieved February 1, 2009, from <http://usmayors.org/bestpractices/litter/NewYork.html>.

This web page from the site of the United States Conference of Mayors was prepared when Mayor Rudolph Giuliani was in office. It presents the origins and purpose of Project Scorecard, a street and sidewalk litter rating system that was one of the first attempts in New York City to measure the results of the city's service delivery. The description provides information about how the measurement is accomplished and provides a table of street and sidewalk cleanliness results in the City's five boroughs in 1976 and 1998.

United States Office of Personnel Management (1974). *Managers Guide for Improving Productivity*: 24.

This report was originally written by Edward Koenig of the Office of Productivity Programs, but the Office of Intergovernmental Personnel Programs revised it for use by state and local governments. It attempts to get managers to look at their organization and its mission, collect and analyze information, make systemic changes, and continue to evaluate and monitor. It incorporates measurement and analysis of inputs, outputs, and efficiency in its methods for increasing productivity. This is an early report in the field of performance measurement, but it lays out the basics.

Williams, M. C. (2005). Can Local Government Comparative Benchmarking Improve Efficiency?: Leveraging Multiple Analytical Techniques to Provide Definitive Answers and Guide Practical Action. L. Douglas Wilder School of Government and Public Affairs. Richmond, VA, Virginia Commonwealth University: 246.

This thesis looks at the attempts at municipal benchmarking in order to discover what benchmarking tactics lead to efficiency improvements. The failed Innovation Group project and the North Carolina benchmarking project are both examined. The study examines the differing ability of municipalities to improve from the information provided from benchmarking and makes conclusions about the effectiveness of voluntary benchmarking.

Willoughby, K. G. (2004). "Performance Measurement and Budget Balancing: State Government Perspective." Public Budgeting & Finance **24**(2): 21-39.

This research presents data from a survey conducted as a component of the multiyear effort by the Governmental Accounting Standards Board (GASB) regarding the Service Efforts and Accomplishments (SEA) research. (p.26) In addition to the case research, the multiyear effort involved a follow-up mail survey conducted in the summer of 2000 to state and local government budget officers and specific agency and department heads and program administrators. This research assesses state personnel perceptions of performance measurement use and effectiveness for both management and budgeting decisions.

Woodbury, K., Brian Dollery, and Prasada Rao (2003). "Is Local Government Efficiency Measurement in Australia Adequate? An Analysis of the Evidence." Public Performance & Management Review **27**(2): 77-91.

Attempts to enhance the efficiency and effectiveness of local government have lagged behind the higher tiers of governance in Australia, and it is only in the comparatively recent past that systematic efforts have been made to measure the performance of Australian local government. This paper reviews municipal efficiency measurement in Australia to advance the argument that the present reliance on partial measures of performance is inadequate and should be heavily augmented by data envelopment analysis (DEA). The authors summarize progress made in efficiency measurement on a state-by-state basis and then examine performance measurement in water and wastewater as a more detailed case study. On the basis of this evidence, the authors argue that DEA provides the best means of providing public policymakers with the necessary information on municipal performance.

Worthington, A. (1999). "Performance Indicators and Efficiency Measurement in Public Libraries." Australian Economic Review **32**(1): 31-42.

A sample of one hundred and sixty-eight New South Wales local government libraries is used to analyze the efficiency measures derived

from the non-parametric technique of data envelopment analysis. Depending upon the assumptions employed, 9.5 per cent of local governments were judged to be overall technically efficient in the provision of library services, 47.6 per cent as pure technically efficient, and 10.1 per cent as scale efficient. The study also analyses the posited linkages between comparative performance indicators, productive performance and non-discretionary environmental factors under these different model formulations.

Worthington, A. C., and Brian E. Dollery (2002). "Incorporating Contextual Information in Public Sector Efficiency Analyses: A Comparative Study of NSW Local Government." Applied Economics **34**(4): 453-464.

Using the planning and regulatory function of 173 NSW local governments, several approaches for incorporating contextual or non-discretionary inputs in data envelopment analysis (DEA) are compared. Non-discretionary inputs (or factors beyond managerial control) in this context include the population growth rate and distribution, the level of development and non-residential building activity, and the proportion of the population from a non-English speaking background. The approaches selected to incorporate these variables include discretionary inputs only, non-discretionary and discretionary inputs treated alike and differently, categorical inputs, 'adjusted' DEA, and 'endogenous' DEA. The results indicate that the efficiency scores of the five approaches that incorporated non-discretionary factors were significantly positively correlated. However, it was also established that the distributions of the efficiency scores and the number of councils assessed as perfectly technically efficient in the six approaches also varied significantly across the sample.

## APPENDIX B

### METHODOLOGY

SPPA employed a methodology for the literature reviews that consisted of thorough searches in all of the major databases representing social and policy journals. In consultations with SPAA faculty, the research associates used the knowledge from the early stages of the research to refine and expand the search, as was appropriate. SPAA created or used the following tools to assure we accomplished the aims of the Commission:

- LUARCC questions from the RFP
- Bibliographic databases
- Search terms specific to each report
- Reference database
- Annotated bibliographies
- Articles/web resources

In the first phase of the basic strategy, SPAA devised appropriate search terms and used them to search all the relevant bibliographic databases. Promising articles were recorded as citations in the reference database (EndNote®) and the full-text article content was saved to the article file (as a PDF).

In the second phase of the search, research associates assigned to the project thoroughly examined the articles for findings that addressed LUARCC's questions. The research associates recorded each finding in the reference database, attaching it to the question it addressed and evaluating the degree to which the article or information resource substantiated the finding. The research associates completed this substantiation, weeding out articles based solely on ideology and rating the remaining articles according to the amount of evidence or the concreteness of the argument presented. During this process, team members reported any new search terms or resources they encountered to the project librarian and project director. Subsequent searches utilized this additional information.

The supervising faculty member assigned to this project and the project director reviewed the results of the first phase of the search, which found the potential articles. When necessary, they redirected the search to clarify information or find new sources. In addition, the faculty members and project director reviewed each finding for relevance to the LUARCC questions.

SPAA sought LUARCC's input through the review of an outline of the proposed methodology to ensure that the literature search was comprehensive and focused on the questions LUARCC raised. Because of the iterative process used in the information review, the early results of our search enhanced the project tools further.

## **LUARCC Questions**

SPAA extracted the questions from the LUARCC RFP into a spreadsheet and distributed it to all team members assigned to the project. The research team coded each of these questions with a Question ID, using the listing in the spreadsheet. This allowed the research associate to link a finding from an article to the question it addressed using the abbreviated Question ID. The questions and Question ID are in the appendix.

## **Bibliographic Databases**

The Bibliographic Database appendix lists all the literature and news databases and other information resources, such as websites and books that SPAA used to find relevant information for LUARCC. The researchers and the project librarian knew certain resources were more comprehensive and relevant for this project. SPAA focused on *Lexis/Nexis Academic* as a resource for periodicals, including newspaper articles, journal articles, and legal documents; *Academic Search Premier* as the best resource for academic journals; *Business Source Premier* as a similar resource; and *Public Affairs Information Service* for government reports, monographs, and articles.

## **Search Terms**

The search within the bibliographic databases and in other resources used the following terms:

- Efficiency measure/indicator/metric
- Productivity measure/indicator/metric
- Cost effective
- Cost-efficiency
- Cost measurement
- Performance benchmarking
- Comparative performance measurement
- Municipal benchmarks
- Performance comparison
- Cost accounting
- Operating cost benchmarking
- Managing for results
- Output cost analysis
- Outcome-based management
- Cost saving
- Municipal efficiency
- Total quality management

The searches always used the following additional governmental terms as limiters:

- Municipality
- Local government
- Borough
- Village
- Town
- Township
- City
- County
- State

Since the search engines differ in their filter capabilities (and the Boolean logic which drives some of them), the research associates used their familiarity with literature searches to make efficient search decisions. They focused on North American articles, cross-cultural studies that included the United States, or other articles, despite their geographic origins, if they were applicable to the subjects the Commission wanted to investigate in a New Jersey context.

Winnowing the search results while casting a broad enough net is not always easy, and never automatic, as our preliminary analysis showed. The research associates, with faculty guidance, used their expertise in the content area they investigated and their experience with literature reviews as doctoral students to make the searches practical and the results comprehensive and responsive.

### **Reference Database – EndNote®**

SPAA summarized the important information from the articles and reports found in the literature searches in EndNote®, a high-quality bibliographic software program by Thompson Reuters. As part of its work products, SPAA will provide a copy of EndNote® to LUARCC with the on-line annotated bibliography for all five projects it completed. A printed copy of the annotated bibliography for this project is Appendix A to this report.

The findings are the core of the reference database and of this project. The findings contain a brief extract from the article or a summary of a key point that links the article to LUARCC's questions. The following information is contained with the findings to provide a quick review:

- **Question ID** – this will contain a project designation and identify the specific question addressed
- **Finding** – a brief description of the finding, conclusion or discussion
- **Cases** – if it is based on an empirical study, the number of cases that support the finding

- **Basis** – reviewer’s evaluation of the substantiation of the finding in the source on a 0 to 4 scale. If there is no basis, it is rated a 0; poor evidence, a 1; some evidence, 2; well substantiated, 3; considered a fact, 4
- **Unit** – the government unit or region cited in the finding

It is not necessary to use EndNote® to make the most of the project completed by SPAA. Rather, the working details contained in EndNote® are there to support this report.

### **Library of Documents**

Each article, or an appropriate extract from the article, report, book excerpt or other research document, has been stored and will be accessible in a separate file as a Word Document or PDF. LUARCC will be able to link to it from the EndNote® reference database that SPAA will provide as a final work product. It will contain the full text of appropriate parts of the information sources found. Hence, it will allow LUARCC or its staff to review the original information that SPAA has summarized, allowing the Commissioners to be confident about their decisions as LUARCC goes forward.

### **Preparing the Final Report**

SPAA used the findings in the EndNote® annotated bibliography to organize the report for each project around the questions initially specified by LUARCC in the original RFP. In preparing the report, the faculty and project director reviewed the findings, interacted with the research associates who had found and read the articles, and, when warranted to resolve any ambiguity, read the original texts. This final review generated SPAA’s conclusions in this report. In addition to this report, SPAA will provide an overall summary report for all five projects it completed.

## APPENDIX C RFP QUESTIONS

RFP Question	Question ID
<b>Through analysis of existing case studies, this work will examine attempts to measure productivity outcomes through the application of a methodology which is credible according to the investigator. The primary purpose is to ascertain whether there are methodologies of benchmarking, performance evaluation or “best practices” for determining the most efficient delivery of municipal services within the United States.</b>	PM
How was efficiency measured?	PM1
Is there any correlation between size of service area for the delivery of a given service and efficiency and/or effectiveness?	PM2
What type of information was collected, and any changes that needed to be made in municipal record keeping to report the data consistently over time?	PM3
How was the information utilized?	PM4
Was the municipality compared to others in a larger geographic study or was the municipality compared only to its own progress over time?	PM5
Did evaluative studies identify cost savings which resulted in property tax savings through greater attention to municipal productivity?	PM6
Were there secondary benefits to the measurement of productivity that were identified in the case studies other than cost savings?	PM7
Were benchmarks established and if so, how was the benchmark utilized?	PM8
Were changes in government structure or other government reforms effectuated as a result of productivity benchmarking?	PM9
Were best practices established and if so, how were they established and how were they used?	PM10
Does the literature point to a model of efficiency measurement we might use in New Jersey to improve municipal service delivery performance?	PM11
What role did the State have, if any, in other States where serious attempts were made at measuring service delivery or was the impetus only from within the municipality itself?	PM12

## APPENDIX D

### BIBLIOGRAPHIC DATABASES

Academic Search Premier  
Books in Print with Reviews  
Business Source Premier  
Catalog of U.S. Government Publications  
Conference Board Research Collection  
CQ Researcher  
Eagleton Poll Archive  
Factiva  
Gartner Research  
GPO Access  
Historical Statistics of the United States: Millennial Edition  
Human Population & Natural Resource Management  
Index to Legal Periodicals Full Text  
Investext Plus  
Lexis/Nexis Academic  
Lexis/Nexis Statistical  
National Technical Information Service  
New York Times  
Public Affairs Information Service - International and Archive  
SA ePublications Social Science & Humanities Collection  
Social Sciences Full Text  
Statistics Canada's E-STAT  
TV-NewsSearch: The Database of the Vanderbilt Television News  
Archive  
Value Line Research Center  
Wilson OmniFile Full Text, Mega Edition  
Worldwide Political Science Abstracts

#### **Other Information Sources**

Research and Information Resources for Public Administration  
[Using the Political Science and Government Literature](#)  
[Public Performance and Measurement Review](#)  
<http://spaa.newark.rutgers.edu/sharedservice.html>  
SPAA's Public Performance and Reporting Network