USE AND SENSEMAKING OF PERFORMANCE MEASUREMENT INFORMATION BY LOCAL GOVERNMENT MANAGERS: THE CASE OF QUEBEC'S MUNICIPAL BENCHMARKING SYSTEM

By Étienne Charbonneau

A dissertation submitted to the
Graduate School-Newark
Rutgers, The State University of New Jersey
in partial fulfillment of requirements
for the degree of
Doctor of Philosophy
School of Public Affairs and Administration
Written under the direction of
Dr. Marc Holzer and approved by

_____

Dr. Marc Holzer

_____

Dr. Norma Riccucci

_____

Dr. Gregg Van Ryzin

_____

Dr. Kaifeng Yang

Newark, New Jersey

May 2010

Abstract

Use and Sensemaking of Performance Measurement Information by Local Government Managers: The Case of Quebec's Municipal Benchmarking System

By Étienne Charbonneau

Dissertation Committee Chair: Dr. Marc Holzer

It has long been taken for granted by researchers that performance data would lead to better informed decisions. Despite the copious literature on performance measurement in the public sector, there is surprisingly very little in the public sector performance measurement literature about what happens once the performance measurement data has been collected. The present research concentrates on (1) the use of performance information, and on (2) a specific facet of use called sensemaking. Akin to interpretation, sensemaking activities are a sub-phase contained between data collection and the taking of actions.

The case of the mandatory benchmarking system covering all 1113 municipalities in the province of Quebec, Canada, is used as a background to expand our knowledge on how managers use and make sense of performance information. Using a mixed-method approach, this dissertation answers two questions. First, what are the factors accounting for the uses of performance measurement by municipal managers? Second, when presented with raw performance measurement information, how do local managers decide if their municipality's performance is satisfactory or not? The quantitative data originate from 312 electronic surveys of General Managers, and official performance values for standardized municipal indicators. The qualitative data takes the form of

content analysis of survey comments from managers, and from focus groups of 179 participants representing 100 municipalities and municipal organizations.

On the use on performance information, the findings from the qualitative data are that managers perceive that external factors hinder their use of performance indicators, and resent being measured and compared. The findings of the quantitative analyses find that attitudes of managers themselves and performance, and not external factors, account of the uses of performance indicators. On sensemaking, the findings are that managers think more in satisficing than in maximizing terms. They are optimistic in their verdicts. Managers do not think of negative and positive performance in consistent terms, but tend to present their performance in the best possible light. The implications of the findings are, in the absence of official definition of what constitutes positive and negative performance, corrective actions following performance information are unlikely.

Acknowledgements

Many people have helped me through the academic chapter of my life. I am taking this opportunity to thank them for their essential support.

First I would like to thank my spouse, Myrna. She is the one that sowed the idea of pursuing a doctoral degree. Five years into our relationship, she agreed to put our union in limbo for the sake of my dreams. Myrna's support was unwavering over 600 kilometers and one border, for four years. Thank you for being there for me; *choukran kteer habibi*[1]!

I want to thank my dissertation chair, Dr. Marc Holzer. Dr. Holzer played many roles during my stay at the School of Public Affairs and Administration. He was a remarkable professor, the Dean of the school, my employer, and my mentor. I learned many things from Dr. Holzer through his teachings, his advice, and his actions. What I appreciate most from him is the confidence and the latitude he bestowed me as a student and a research assistant. I would also like to thank Dr. Norma Riccucci and Dr. Gregg Van Ryzin. In addition to their counsel for this dissertation, they provided me with guidance in the doctoral program. Both professors were generous with their time. Dr. Riccucci's ideas enriched the way I see the world; Dr. Van Ryzin graciously and consistently shared with me habits and skills that he had cultivated as a scholar. These three professors are epitomes of what scholars and mentors ought to be. I also want to thank Dr. Kaifeng Yang, of Florida State University's Askew School of Public Administration & Policy, for his comments and support.

[1] *Thank you a lot my love!*

My colleagues and friends at SPAA deserve an acknowledgement. Daniel, Morgan, Alexander, Ashley, and Mark offered suggestions and comments on my dissertation. All of you proved to be great sounding boards. Gautam, Aroon, and Alicia also deserve my gratitude for the coaching they offered to me early on in the program. I want to thank the eight of you, and the student computer lab crew, for making Newark a friendly place.

I also want to thank Dr. François Bellavance, Stéphane Messier, and Dr. Michel Guindon at HEC Montéal, and Jean Monfet and Claude Brochet at MAMROT, for their support in the data collection stage of this project. I want to thank Julie Faucher and the active members of the *Partners on Municipal Management Indicators Committee* for their invaluable help. Without the many reminders they sent to municipal managers, the response rate to my survey would have suffered intolerably.

Finally, I want to thank my mother, Lucie, and my brother, Guillaume. It has been a long and improbable journey from Saint-Esprit, QC, to Rutgers University. I would not have made it without you. *Merci beaucoup*[2].

[2] *Thanks a lot.*

Table of Contents

List of Tables

List of Illustrations

Chapter 1: Introduction

In 1896, Willcox wrote that "no practical object would be gained by a discussion of the economic productivity of roads or parks or sewers or police. They all have an economic side, and might be justified from that standpoint. But it is needless to justify what everyone accepts" (Willcox, 1896:378). More than a century later, a great deal of discussion has occurred among politicians, managers, corporate and citizen interest groups, and scholars to justify the performance of governmental services. Performance measures have been foci of much debate, and the bases on which promises of efficiency and reduced cost have been made (Rogers, 2006:45).

A large body of theoretical research on performance measurement in the public sector examines the determinants of its existence (Streib and Poister, 1999; Holzer and Yang, 2004; Chung, 2005), its implementation (Palmer, 1993; Johnsen, 1999; Lawton et al., 2000; De Lancer-Julnes and Holzer, 2001; Ho and Chan, 2002; Hill and Andrews, 2005; Jordan and Hackbart, 2005), its perceived benefits (McGowan and Poister, 1985; Berman and Wang, 2000; Rogers, 2006) and shortcomings (Radin, 1998; Halachmi, 2002b, 2005; Hood, 2007). Despite the copious literature in performance measurement studies, there is surprisingly very little in the public sector performance measurement literature about what happens once the performance measurement data has been collected. It has long been taken for granted by researchers that from the presence of performance data would come (better) informed decisions. As a result, "while the production of performance information has received considerable attention in the public sector performance

measurement and management literature, actual use of this information has traditionally not been very high on the research agenda" (Van De Walle and Van Dooren, 2008:2).

The present research will concentrate on a specific facet of performance information use called sensemaking. Akin to interpretation, sensemaking activities are a sub-phase contained between data collection and the taking of actions in Daft and Weick's (1984:286) general framework. This interpretative step is part of the information utilization phase involved in the overall performance system framework (Lu, 2008:12). In simpler terms, sometime between receiving performance information and taking actions in reaction to this information, managers have to come up with a verdict about what the information means. Palmer and Short (2001) summarize this reality that also exists in the private sector. They state that "performance does not come tidily packaged and labeled for executives to act upon. Instead, it must be interpreted through referents before actions are chosen" (Palmer and Short, 2001:211).

Framework

For the purpose of this research, Jeong and Brower's (2008) *Organizational Sensemaking Activity in Three Stages and Three Contexts* framework will circumscribe the study of sensemaking. More specifically, sensemaking in this research will be limited to the second stage of the above mentioned framework.

Table 1: Organizational Sensemaking Activity in Three Stages and Three Contexts

Context	Stage		
	Noticing	Interpretation	Action
Ecological context	Capturing some problematic portion of the field of experiencing the working situation outside	Making the symbolic reality of the captured portion of the field of experience	Externalizing the symbolic reality made in the mind, and in doing so shaping the contour of the working situation outside
Institutional context	Questioning the preorganized scheme of reference in the mind	Reorganizing the scheme of reference in the mind	Actualizing the reorganized scheme of reference, and in doing so channelizing the subsequent organizing process into either thematic or prethematic mode
Social relational context	Revealing my subjectivity to the mind, which is incompatible with your subjectivity	Establishing my new subjectivity; referring to your subjectivity.	Objectifying my subjectivity, and in doing so making it available to us, involving me and you, as elements of a common world

Source: Jeong and Brower (2008:243)

The case of the mandatory benchmarking system covering all 1113 municipalities in the province of Quebec, Canada, will be used here to further our knowledge on how managers make sense of performance information.

Why the topic matters

The topic of this dissertation matters because it addresses a neglected fundamental issue in the measurement of governmental services: the lack of calibration. After much data has been collected, managers trying to make sense of indicators' values have little guidance to assess if the performance of their agency is high or not (Meyer and Zucker, 1989:25). This problem of distinguishing poor and good performance, although present for agencies offering direct services, like municipal water and waste treatment services, is

compounded with agencies responsible for delivering less tangible services like social care (Clarkson and Challis, 2006:462). It is up to managers to decide if the recorded performance is a flatout failure or a success, or whether it is satisfactory or not. To that extent, Kenis (2006) suggested, but did not test, a relevant question capturing the essence of sensemaking:

> (...) which criteria do actors (whether they be individuals, groups or organizations) use to assess themselves, what type of information do they gather on how they are doing, and how do they use this information to eventually adjust to deviations; or in a somewhat shortened version: 'how do actors know how they are doing, and what do they do with the information about how they are doing? (Kenis, 2006:117)

Knowing how managers make sense of information to arrive at a verdict about performance informs us about the origins of governmental actions. For example, in the private sector, a verdict of failure is "(…) principally conceived in terms of organizational scale (employees, turnover, product range) or market share, criteria that do not translate into the public-sector environment where scale is a function of legal and political decisions, and services still retain a quasi-monopolistic status" (Jas and Skelcher, 2005:199). There are real costs in the misinterpretation of government performance.

> There are costs to the public if PM [performance measurement] fails to identify under-performing units so that no remedial action is taken. There are real but less-well-recognized costs of falsely identifying a unit as underperforming when in fact there is no significant difference between it and others judged as 'adequately performing'. There are effects on staff, staff morale and recruitment which in turn impact negatively on the service provided. Extensive management and organizational changes, if triggered by false identification, have disruption costs that will not be justified by improved performance in the longer term. (Bird *et al.*, 2005:20-21)

Repercussion of sensemaking can be sensed at the macro and micro levels of an organization. At the employee level, the argument is that there is a need to have a basis of

comparison: "(…) if a person does not have any of this type of information, he or she will have difficulties in performing the job" (Kenis, 2006:122). At the level of organizational strategy, it was found that for private organizations sensemaking activities affect strategic course of action.

> Managers of firms achieving favorable levels of performance, relative to their referents, are likely to persist with the status quo because they employ exploitation of the known, or low-level learning. On the other hand, performance levels which are deemed unsatisfactory relative to their chosen referents are more likely to instigate a problemistic search (…) and exploration of the new, resulting in altered strategies. (Palmer and Short, 2001:212)

Other scholars/researchers have considered the salience of the need of making a normative judgment on the achieved level of performance in order to make performance information actionable by users. Hatry (2008:226-227) identified it as one of the "five key 'technical' elements that seem necessary for successful use of performance measurement information."

The main interrogation of this research, which is also its potential contribution to the field, goes as follows: When presented with raw performance measurement information, how do local managers decide if their municipality's performance is satisfactory or not? It was hinted, but not answered, by Raaum (2007). Discussing the role of auditors, Raaum affirmed that:

> The relevant question here is, how do we auditors determine the operations cause-whether work is being done well or poorly? (…) Let's hope we don't just guess. (…) To do this, it seems to me that auditors need some standard or benchmark of how work for a given operation should be done. (Raaum, 2007:49).

More specifically, this research attempts to answer the following questions:

R1. Which factors account for the uses of performance measurement by municipal managers?

R2. What are the comparison levels being used by municipal managers in interpreting performance measures?

R3. How are targets set by municipal managers?

R4. While interpreting their own performance, when do municipal managers deem that the municipality's performance is (a) a failure, (b) unsatisfying, (c) satisfying, (d) a success?

The next chapter will review the literature on performance measurement, benchmarking, use of information, and sensemaking. Chapter 2 presents the context of the Municipal Management Indicators benchmarking system in Quebec. Chapter 3 explains the methodology used, that is survey research and regression analyses. Chapter 4 shows and discusses the results of the quantitative analyses. Chapter 5 shows and discusses the results of the qualitative analyses. Chapter 6 offers conclusive remarks.

Chapter 2: Literature Review

Performance Measurement

Basics - performance measurement

Performance measurement is a pragmatic information-based management tool. In their performance measurement activities, managers collect information on a predetermined limited list number of administrative activities and their alleged results. The idea behind performance measurement is to simplify the administrative reality into a few graspable dimensions. This positivist ontological position favors a shared structured set of organizational goals over a heteroclite myriad of personal goals. Whether a measure is a true and fair representation of what is being measured is an "ontological issue that is not at the core of the pragmatism characterizing management" (Catasùs *et al.,* 2007:508). This pragmatism has been taken as far as to state that "how a measure is used is more important than what the measure is. There is no such thing as a bad performance indicator, only bad use of performance indicators" (MacPherson, 2001:17). This however, is not an orthodox view on performance measurement.

Many definitions have been offered for performance measurement. It has been described as the "(…) now established practice of the use of performance indicators in the public sector to represent a "picture" of performance to an external constituency" (Ball, Bowerman and Hawksworth, 2000:23). The timely flux of information is key to understand performance measurement, since performance measurement is the "regular and careful monitoring of program implementation and outcomes. This regularity is one of the characteristics of performance measurement that differentiates it from program evaluation" (de Lancer-Julnes, 2006:223). A more comprehensive definition by Holzer

and Yang (2004: 16) described performance measurement as an "(…) opportunity to present evidence that the public sector is a public bargain, to highlight the routine but important services that public servants quietly provide and to answer the public's sometimes angry questions and implicit suggestions on a dispassionate basis." The integration of performance information into management is the essence of performance management, which is a "(…) system that generates performance information through strategic planning and performance measurement routines and that connects this information to decision venues, where, ideally, the information influences a range of possible decisions" (Moynihan, 2008b:5). Performance management would consist of three functions: measurement, analysis, and communication (Askim, Johnsen and Christophersen, 2008:299). These functions resonate with the three broad aims of public service performance data, as identified by Bird *et al.* (2005:2): establish 'what works' by promoting stated objectives, identifying the functionalities of individual practitioners or organizations, and reporting on the stewardship of the public services.

The availability of information about public services delivery was sought to limit demagoguery and elevate the public debate on governmental operations, which in turn would be an incentive to improve the performance of the services (Hamilton, 1972:795). Since government agencies often find themselves in monopolistic positions, performance indicators would carry information, "(…) functioning as 'prices' in political markets, in much the same way as prices do in input and product markets" (Johnsen, 2005:14). Once indicators are determined, information about these indicators can be collected from variety of sources like citizen surveys, employee time sheets, official records (financial and non-financial), etc. (Foltin, 1999:43). Strictly speaking, "a *performance indicator*

implies a signal whereas a *performance measure* implies a more scientific technique involving comparison to a yardstick or target" (Collier, 2006:166). In the practitioners and academic literature, the two terms are used interchangeably. As a case in point, the English *Public Administration Select Committee* (2003:5) defined performance indicators as "quantifiable measures used to monitor performance and report it to the public." From this definition, one might notice that performance indicators are really constructions; they are not truly neutral signs of the consequences of governmental interventions (Frønes, 2006:20). As such, they are pragmatic attempts to represent reality by offering one suggestive definition instead of letting many suggestive definitions float around.

The rationale often put forward for managers to explain why they should have access to timely and reliable performance measurement is coined in the famous aphorism 'what gets measured gets done'. Halachmi (2002b) offers an 'incomplete list' building on this intuitive idea:

1. If you cannot measure it you do not understand it;
2. If you cannot understand it you cannot control it;
3. If you cannot control it you cannot improve it;
4. If they know you intend to measure it, they will get it done;
5. If you do not measure results, you cannot tell success from failure;
6. If you cannot see success, you cannot reward it;
7. If you cannot reward success, you are probably rewarding failure;
8. If you will not recognize success you may not be able to sustain it;
9. If you cannot see success/failure, you cannot learn from it;
10. If you cannot recognize failure, you will repeat old mistakes and keep wasting resources;
11. If you cannot relate results to consumed resources you do not know what is the real cost;

12. If you do not know the actual cost you cannot tell whether or not you should do or outsource it;
13. If you cannot tell the full cost you cannot get the best value for money when contracting out;
14. If you cannot demonstrate results, you may undermine your ability to communicate with important stakeholders and you cannot win public support because you do not provide value for money. (Halachmi, 2002b:65)

It should be noted that this appeal to common sense is found more in practitioner publications and policy think tanks literature than in the academic literature. With the notable exception of Catasùs *et al.* (2007), these "Gulickian" proverbs are seldom studied empirically. To that extent, despite decades of research, "(…) we still do not have a good understanding of the performance measurement-effectiveness relationship" (Mausolff and Spence, 2008:596).

The following section will present the perceived benefits of performance measurement in the public sector.

Benefits of performance measurement

Performance measurement and management are being pursued internationally because they are seen as being potentially advantageous to service delivery. For methodological reasons akin to the ones presented in the previous section, it is difficult to empirically attribute changes in organizations to their performance measurement activities. When commenting on the Citistat model in Baltimore, MD, Behn (2008) crystallized this methodological impediment of empirically demonstrating the positive impacts of performance measurement.

> Q: Can Baltimore "prove" that CitiStat was the cause [of performance improvement]?
> A: Of course not. Any change in the results produced by a public agency has many causes. Rarely does a public agency take only one action while carefully holding the rest of its behavior faithfully constant. And even if, to examine the impact of this one action, the agency tried and was able to do so for a long enough period of time (years? decades?), there would still exist a variety of external factors that are constantly changing and which do—or, at least, might—have an impact on the results. And it is difficult to rule out any impact from many of these potential causes. After all, for any improvement in any kind of performance, one possible explanation is always regression towards the mean. (Behn, 2008:43)

Other approaches have been used to support claims about the benefits of performance measurement on public sector performance.

One of the sources of claimed benefits of performance measurement comes from deductive reasoning, often on the part of consultants. One such example comes from Page and Malinowski (2004), when they 'find' that one of the benefits of performance measurement is that it helps public agencies to monitor, penalize, and reward contractors. According to them, "good measurement systems help break the cycle of low-bid/low-quality procurement by rewarding high-performing contractors with contractor extensions and other preferences" (Page and Malinowski, 2004:32). Benefits like 'overtly prioritizing tasks' have also been uncovered by deductive reasoning on the part of practitioners, for example Chief Police Inspectors (Rogerson, 1995:26). From the public sector performance literature, De Bruijin (2002) finds three recurrent benefits of performance measurement that are identified this way: (1) Performance measurement brings transparency; (2) Performance measurement is an incentive for output; (3) Performance measurement is an elegant way of shaping accountability (De Bruijin, 2002:580-581). Without discrediting the bases used to identify the benefits of

performance measurement altogether, one has to acknowledge that benefits found with deductive reasoning are more convincing to performance measurement proponents than opponents.

A second source of performance measurement benefits originates from inductive reasoning, often from single case studies. For example, from the cases of Larvik, Norway and Austin, TX, we learned that performance measurement can respectively foster active scanning for resources (Askim, 2004:434) and help make targeted budget cuts in economic downturns (Plant and Douglas, 2006:47). The benefits identified in that fashion are primarily found in professional magazines and programs evaluation in governmental reports. Taken together, the sheer number of such studies crediting performance improvements to performance measurement activities is impressive. However, individually, these findings are less convincing to the empirically biased, and one could say to the ontologically obtuse, researchers.

A third source for identifying benefits of performance measurement is systematic practitioners' surveys and interviews. These surveys and interviews tapped the perception of managers using performance measurement in their activities. Despite their subjectivity, the benefits identified obtained through wide systematic samples of practitioners offer more convincing findings than the ones from generic deductive reasoning or single case-studies inductive reasoning. One early study using 460 of the 1,062 ICMA local managers from municipal jurisdictions with between twenty thousand and one million citizens was done by McGowan and Poister (1985). They found that managers felt that performance measurement had some influence in "(…) modifying program objectives

(51%), altering work standards (53%), and setting individual performance targets (49%)" (McGowan and Poister, 1985:537). Additionally, they found that 61% of respondents found that the benefits of performance measurement outweigh the cost of collecting the data; less than 1% felt it was not (McGowan and Poister, 1985:537). Another study using later ICMA data by Poister and Streib (1999) became a classic article on the benefits of performance measurement. To this day, despite the facts that the study is based on a small number of usable surveys for their purpose (municipalities actually using performance measurement) and that it is a decade old, it is still the most cited study on the benefits of performance measurement. Overall, Poister and Steib (1999:331) found that most managers indicated that their performance measures improve decisions at least moderately and reported at least moderate changes in budget allocations. Berman and Wang (2000), using 209 surveys from county managers in the United States, identified the extent to which managers found the following benefits to be present in their county as a result of the use of performance measurement. Overall, performance measurement received limited credits for program outcomes' variations. Managers 'agree or strongly agree' with the potential benefits of performance measurement in the following proportions:

1. Increased awareness about the need for accountability (48%);
2. Increased ability to determine service efficiency (45%);
3. Increased ability to determine service effectiveness (43%);
4. Increased ability to determine service timeliness (40%);
5. Established performance target levels for programs/services (40%);
6. Clarified agency or program goals and objectives (37.2%);
7. Improved accountability of program performance (35.6%);
8. Ability to achieve improvements despite resource constraints (32.5%);

9. Increased commitment to excellence (31.5%);
10. Improving group decision-making capabilities (26.6%);
11. Determined long-term budget needs (23.6%);
12. Elimination of no longer needed services (16.1%);
13. Improved timeliness of management decisions (15.5%) (Berman and Wang, 2000:417).

From interviews in Portuguese fire service administration, Carvalho and colleagues (2006:174) found that performance measurement would have the following benefits: "supporting the process of planning and developing budgets; contributing to the accountability of services; increasing the motivation of fire-fighters through systems of incentives, rewards and sanctions; and stimulating the citizens' interest in these public services." Also relying on interviews, this time in the Department of Correction in Alabama, Moynihan (2005:229) found that performance measurement, through 'managing for results' reforms, was not perceived to bring improvements in allocation, effectiveness, and efficiency. However, the benefits would be more symbolic, showing that managers of prisons cared about performance (Moynihan, 2005:229).

All in all, performance measurement is widely credited with improving performance and increasing accountability. The proofs to such claims in the literature vary in methodological robustness. Proponents of performance measurement will be reassured by the abundance of consultants and practitioners concluding that performance measurement helped them reduce cost and communicate better with their constituents. They will also cite some studies where a vast number of managers identify perceived benefits of performance management. Opponents of performance measurement will point out that an irrefutable empirical demonstration of the benefits of performance measurement in the

public sector has yet to come. In the following section, the barriers of performance measurement will be discussed.

Barriers to performance measurement

For some seasoned managers who have seen the coming and going of management tools like *management by objectives*, *zero-based budgeting*, *performance-based budgeting*, *total quality management*, the *balanced scorecard* and the *organizational dashboards,* waiting out on management tools until they go away seems like a sensible strategy (Behn, 2008:9). Understanding the perceived barriers against performance measurement in the public sector will help us comprehend the mitigating factors in the use of performance measurement. Reviewing the barriers of performance measurement is different from enumerating its perceived shortcomings. The perceived shortcomings of performance measurement will be presented in another section. Additionally, Smith's (1995:283) famous unintended consequences of performance measurement will not be covered here. Gaming, the most perverse potential incentive of performance measurement, will be covered in its own section. In this section, results from a meta-analysis and national empirical municipal studies done in Canada, the United States, the United Kingdom, Finland and Sweden will be presented in turn.

Johnsen (2005) tries to draw lessons from the last twenty-five years of the performance measurement literature in the public sector. Choosing from articles using deductive reasoning, inductive reasoning and empirical studies, he presents an inventory of the obstacles facing performance measurement. He classifies these obstacles in three categories: obstacles to implementation, obstacles to use, and obstacles to outcomes. The

obstacles to implementation and use are worth representing here; the obstacles to outcomes are similar to the ones in Smith's (1995) study. Johnsen (2005:11) lists seven obstacles often found in the literature in regards to the implementation of performance measurement. They are (1) low theoretical or methodological planning; (2) competence ('data-driven' implementation); (3) small resources/capacity for development; (4) inadequate implementation structures; (5) low top-management commitment; (6) impatience; (7) resistance (Johnsen, 2005:11). He also lists twelve obstacles identified as limiting the use of performance measurement in government: (1) lack of relevant statistics and data; (2) measurement errors; (3) misinterpretation; (4) low decision relevance; (5) running down of PIs [performance indicators]; (6) proliferation of PIs; (7) information overload; (8) indirect lines of responsibility (common agency); (9) no ownership of performance; (10) loyalty to professional norms rather than to management; (11) misrepresentation (manipulation of data); (12) 'creaming' (Johnsen, 2005:11). Johnsen's lists are interesting, but little is explained about how these lists were constructed. The bibliography in his article certainly does not cover all that is relevant in the twenty-five years preceding the publication of his article.

The five studies to follow represent empirical research relying on surveys and interviews of municipal managers. Readers interested in their respective methodologies should consult the articles in question. Here is a summary of their findings concerning the barriers perceived by municipal managers.

In a North American context, Pollanen (2005) analyzed 334 surveys from senior administrators in municipalities with populations greater than 5,000 in five Canadian provinces (British Columbia, Saskatchewan, Ontario, Nova Scotia, and Newfoundland). The author found that the three top-rated factors limiting the development, use, and reporting of performance measures are: (1) the difficulty in identifying appropriate measures; (2) the difficulty in meaningful use of measures, and (3) the ambiguity of performance objectives (Pollanen, 2005:17). In Pollanen's article (2005), another North American study is often referred to: Poister and Streib's (1999) inquiry of the state of performance measurement in the United States, based on 159 municipal managers using a jurisdiction-wide performance measurement system. Some of the problems with performance measures studied by Poister and Streib (1999) resonate with Pollanen's limiting factors, like the timeliness of data collection, use and reporting (Poister and Streib, 1999:332).

Drawing from exploratory surveys, focus groups of managers and semi-structured interviews in the United Kingdom, McAdam and O'Neill (2002) unearthed perceived barriers in the use of performance measurement by managers. Recurring barriers emanating from the interviews were that:

1. Managers are aware of inconsistencies in measurement in relation to a number of indicators;
2. Completion of time sheets in particular is not given priority in some areas;
3. There is an urgent need to educate managers and staff in the interpretation of relevant information from performance measurement;

4. Benchmarking clusters, set up under best value and involving clusters of similarly sized councils, have to date merely agreed to compare with one another, using the (minimum) key performance indicators; and
5. There is a need to set a realistic level of service, taking into account resources available. (McAdam and O'Neill, 2002:452-453)

Interestingly, many of the perceived barriers have to do with a perceived lack of complexity on the part of the system prior to the Comprehensive Performance Assessment (CPA) system. The then upcoming CPA system was arguably the most sophisticated mandatory municipal benchmarking system in the world.

Rantanen and colleagues (2007:422) used multiple-case studies of (1) a Finnish university; (2) a state agency working under the Ministry of Trade and Industry, serving also the Ministries of the Environment, the Interior, and Transport and Communications; and (3) the maintenance function of the Finnish Defense Forces, to study the problems facing Finnish performance measurement systems. From Finnish cases, the authors came up with a classification of the barriers in the design and implementation of performance measurement systems.

Table 2. Classification of the Underlying Reasons and Practical Problems in the Performance Measurement of Public Sector Organizations

	Characteristics of public sector organizations (underlying reasons)	Problems in performance measurement (outcomes caused by the reasons)
Factors affecting the design of a PMS:	Many stakeholders with conflicting needs	Difficulties in solving the conflicts between the needs of different stakeholders (i.e. not clear what should be measured)
	Undefined end products and goals (efficiency vs. effectiveness)	Difficulties in target setting (i.e. not clear what the goal of the operations should be)
Factors affecting the implementation of a PMS:	Lack of ownership of the property	Representatives of different stakeholder groups influence the development of individual measures on a too detailed level
		The personnel does not understand the objectives of the measure development
	Poor management skills	Too many responsible persons in the measurement development lead to non-responsibility
		The personnel does not see the usefulness of the project for their work and ignore or resist it
		Overlapping projects hamper the measurement project because they take resources

Source: Rantanen *et al.* (2007:428)

Again, what the authors labeled as 'problems' in performance measurement (the right column in table 2), are related to the use of performance measures in managerial activities.

In a questionnaire to finance directors or equivalent staff members in all 290 Swedish municipalities, Siverbo and Johansson (2006) asked municipal finance directors to express themselves on the obstacles to the implementation and use of RPE (relative performance evaluation). The list is organized in three categories: obstacles related to how managers or organizations are (1) unwilling or (2) unable to use performance

measurement in their activities or (3) just prevent using it. Under the label "unwilling", the reasons presented are (1) RPE is considered a poor method, (2) users have a (not sure but we need a verb) low trust in ratios, (3) RPE is felt to convey an incomplete picture of the organization, (4) there is a fear of RPE being misunderstood and misinterpreted (Siverbo and Johansson, 2006:283). By "unable", the authors meant that municipalities would have a (1) lack of knowledge about RPE as a possible model, (2) lack of relevant comparison municipalities, (3) lack of available working time, and (4) lack of expertise for working with RPE (Siverbo and Johansson, 2006:283). At last, under the label "prevented", there are obstacles like (1) the municipality has an explicit or implicit policy against RPE, (2) our officials are uninterested, (3) RPE is seen as a threat, (4) there is a fear of exposing possible weaknesses (Siverbo and Johansson, 2006:284)[1].

To finish this section on the barriers to the use of performance measurement, it is worth presenting Ammon's (1985) thirty-seven common barriers to productivity improvement in local government. Keeping these barriers in mind will help to keep the expectations on what performance measurement can realistically achieve in municipal government. It also explores the boundaries on performance in municipal government, by illustrating the topics of upcoming sections in this research, among them complexity and non-linearity. From fourteen case studies throughout the United States, the thirty-seven common barriers to productivity improvement in local government, as described by Ammons (1985), are:

[1] The survey results are not replicated here. Readers interested in knowing the extent in which Swedish municipal managers perceived these to be obstacles in their municipalities are invited to look at the article itself.

1. political factors that influence decision making;
2. productivity improvement's lack of political appeal;
3. short time horizon of politicians and top executives;
4. policy rather than performance emphasis in local affairs;
5. public perceptions regarding changes and benefits;
6. fragmentation of local government;
7. inadequate research, development, and experimentation;
8. antiproductivity effect of federal grant provisions;
9. intergovernmental mandating of local expenditures;
10. civil service restrictions;
11. legal restrictions to motivational programs;
12. barriers to monetary incentive plans;
13. dominant preference for the status quo;
14. absence of market pressures;
15. perceived threat to job security;
16. union resistance;
17. bureaucratic socialization processes;
18. primary emphasis on effectiveness rather than efficiency;
19. lack of accountability;
20. risk avoidance;
21. perverse reward systems;
22. absence of personal rewards for innovation and productivity;
23. conceptual confusion;
24. managerial alibis;
25. inadequate management commitment to productivity;
26. reluctance to abandon;
27. ambiguous objectives and lack of performance measurement;
28. absence of cost accounting systems;
29. inadequate information on intracity and intercity performance;
30. inadequate information dissemination and reluctance to use what is known;
31. inadequate performance evaluation;
32. insufficient analytic skills or analytic staffing;
33. performance myths;
34. requirement of large initial investment for productivity efforts;

35. overselling productivity improvement programs;
36. bureaucratic rigidities and fragmented authority;
37. supervisory resistance.

(Ammons, 1985:295)

The goal of this section was not to explain each and every one of the possible barriers to performance measurement. Rather, it was to offer the reader a panorama of encountered barriers perceived by local managers around the world.

Types of performance measures

There are different types of performance measures because government pursues many goals at once. The nature of the performance measures used in a particular government agency, assuming it did not have measures imposed by a higher level of government, will mirror the goals of that agency. Early in the modern era of the performance measurement literature, Hatry (1980) identified eleven kinds of performance measures. There are (1) cost measures, (2) workload-accomplished measures, (3) effectiveness/quality measures, (4) efficiency/productivity measures, (5) actual unit-cost to workload standard ratios, (6) efficiency measures and effectiveness quality, (7) resource-utilization measures, (8) productivity indices, (9) pseudo-measures, (10) cost-benefit ratios, (11) comprehensive performance measurement (Hatry, 1980:315-322). It should be noted that most of these categories are no longer used thirty years later. A more recent exercise by Willoughby (2004:28) identified eight categories of performance measures: (1) inputs, (2) process/activity, (3) outputs, (4) outcomes, (5) cost/efficiency, (6) quality/customer satisfaction, (7) explanatory, (8) benchmarks. In practice, both in governmental agencies and in the academic literature, the list of types of performance measures is shorter. Two

parallel categorizations are in usage. One recurrent category spans from *inputs*, *outputs* and *outcomes*. The other common typology categorizes measures according to *effectiveness* and *efficiency*. Many definitions exist for these terms. Offering an exhaustive repertory of definitions is beyond the scope of this research. Nevertheless, here are some definitions to guide the reader through the rest of this research.

Inputs are the resources financial and nonfinancial used by an organization, when providing services (Public Administration Select Committee, 2003:5; Willoughby, 2004:28). Outputs are the services, goods or products provided by the organization with the inputs, while taking into account certain quality requirements (PASC, 2003:5; Willoughby, 2004:28). Outcomes are the benefits or values that occur, at least in part, because of services provided (PASC, 2003:5; Willoughby, 2004:28). It was reported that for municipal governments in Canada, outputs are more reported than outcomes: this can be explained by the fact that outcomes are more ambiguous than outputs, thus they are more difficult to measure (Pollanen, 2005:18). Outcomes are more vaporous in the public sector because outcomes would take time to materialize and because results from an unknown combination of outside influences (often socioeconomic forces) cannot be controlled by agencies (Askim, 2004:429). Prescriptions from the literature in are that outcomes should be prioritized over outputs (Edwards and Thomas, 2005:374). This is not to say that no one defends the utility of outputs (Frederickson and Frederickson, 2006:170-171). Another reason why outputs are often measured in the public sector is that the measurement of outputs is rampant in the private sector. In the private sector, measuring characteristics or counting activities is not as much a concern, since the

"market price of the output measures the consumers' marginal valuation of the bundle of characteristics from consuming the output" (Castelli *et al*,. 2007:106). Also, in the public sector, "there are no prices to reveal patients' marginal valuations of services so we have to find other means of estimating their value and we define the quality of the output as a function of the vector of outcomes it produces" (Castelli *et al*,. 2007:106).

Effectiveness refers to "the impacts and quality of the service delivery, whether the service achieves its purpose, and how responsive it is to community needs" (Hatry, 1980: 312). Efficiency is commonly defined as the amount of input (usually monetary expenditures or amount of employee time) required for the amount of output produced (Hatry, 1976:23; Willoughby, 2004:28; Hatry, 2006:7). Contrary to effectiveness, there are many different kinds of efficiency. The most primitive kinds of efficiency used by government are described by Hatry (1976):

> The first and most familiar is unit cost (...) A second type of efficiency measurement is one in which an effectiveness measure is used as the measure of output in the ratio of output to input. This is seldom used form of measurement probably because of the dearth of effectiveness information. (...) The third type of efficiency measure does not use the traditional output to input ratio. This third form measures the degree of utilization of facilities, equipment or employees. (Hatry, 1976:22-23)

Despite the fact that unit costs indicators have been discredited as relics of the industrial age, better suited for mass production of identical units (Guven-Uslu and Conrad, 2008:240), cost unit indicators often come with a denominator of population, as in per capita ratios. The crudeness of these ratios has been denounced in general in the public

sector (Halachmi, 2005:259), questioning if cost ratios should be called efficiency measures at all, and calling attention to the fact that they are "plagued by cost accounting variations from city to city" (Ammons and Rivenbark, 2008:310). Specific rebuttal of cost measures has been done for traffic and road safety (Bailey and Hewson, 2004:502), fiscal health of municipalities (Hendrick, 2004:91), fire services (Ridley, 1927:22), police services (Ridley, 1927:31), and small hospitals (Lied, 2001:168).

In response to naïve efficiency (that is crude cost ratios), more precise measures decomposing the overall efficiency between *allocative efficiency* (the reallocation of resources), *technical efficiency* ("when it is not possible to decrease any input without decreasing outputs or increasing another input and when it is not possible to increase any output without increasing an input or decreasing another output" (Duffy, Fitzsimmons and Jain, 2006:234)) and *scale efficiency* (returns-to-scale) (*see respectively* Cooper, Seiford and Tone, 2006:149,45,142) have been developed in the 1950s (Koopmans, 1951; Farrell, 1957). To this day, these are seldom used (with the notable exception of Finnish municipalities, see Johnsen and Vakkuri, 2006:296) and advocated (with the notable exception of a HM Treasury report on police services, see Spottiswoode, 2000) in the measurement of efficiency by public agencies, but are ubiquitous in the academic literature on public agencies (Chalos and Cherian, 1995; Thanassoulis, 1995; DeBorger and Kerstens, 1996; Kerr *et al.*, 1999; Carrington *et al.*, 1997; Drake and Simper, 2001, 2002, 2003; Grosskopf *et al.*, 2001; Pina and Torres, 2001; Diez-Ticio and Mancebón, 2002; Mante and O'Brien, 2002; Sun, 2002; Worthington and Dollery, 2002; Hougaard, Kronborg and Overgård, 2004; Bretschneider, Marc-Aurele and Wu, 2005; Ouellette and

Vierstraete, 2005; Stevens, 2005; Williams, 2005; Clarkson and Challis, 2006; García-Sánchez, 2006; Barros, 2007; Afonso and Fernandes, 2008; Cordero-Ferrera, Pedraja-Chaparro and Salinas-Jiménez, 2008; Hauner and Kyobe, 2008; Geys and Moesen, 2009; Sarrico and Rosa, 2009).

Concluding on the definitions of the different types of performance measures, two elements should be present to mind. First, the categorization of indicators into input/output/outcomes and effectiveness/efficiency is not an exact science. The same indicator can be categorized differently by different organizations, depending on their mission (Hatry, 2006:13-14). Second, the different types of indicators are not independent but are dynamically linked together. Drake and Simper (2003), in the context of studying police services, described the linking of the types of measures:

> (...) it is important to note that there is a potential circularity problem in any analysis of police force efficiency if: inputs are proxied by costs; costs are closely related to funding; funding is based on the funding formula, which allocates police funds on the basis of perceived need (i.e. predictions of the level of crime); and finally, if the outputs specified contain variables such as the level of crime. (Drake and Simper, 2003:708)

Aside from organizational goals, the reliance on certain types of measures rather than on others reflects the preferences of who short listed the measures.

Preferences of types of performance measures

Preferences for types of performance measures would be important primarily for the use of performance information, but also for sensemaking. Fundamentally, different types of managers have different preferences on what they think should be measured. In turn, the perceptions of which indicators carry over actionable information varies according to ranks and functions in an organization. For example, it has been recognized that

municipal managers in finance and budgeting departments do not report their activities the same way other managers do (Marcuccio and Steccolini, 2009:160). Similarly, marked differences exist between municipal managers in finance and budgeting and other managers regarding perceptions of dysfunctions of performance indicators (Willoughby, 2004:36).

Output measures would be preferred by finance and budget officers at the local level, whereas other managers would be more interested by outcomes. From interviews in five municipalities in south Florida and postal survey of local officials in all 67 general-purpose municipalities with 2,500+ population in Dade, Broward, and Palm Beach Counties, FL, Wang (2002) witnessed "measurement conservatism" on the part of finance directors. They were more likely to select output measures than other groups (Wang, 2002:814). Preferences according to ranks in the municipal hierarchy have been studied. Looking at the state of the practice of municipal performance in England before the adoption of mandatory benchmarking systems, Palmer (1993) found that middle managers had a distinctive preference for output measures, especially cost measures. Her findings are that in general, the concerns of middle management are focused on process, rather than on the effectiveness of service delivery. Thus, the types of performance indicators seen as most relevant include “measures of work performed and the amount of work done in a defined length of time. Again, the focus on cost is illustrated by the fact that 93% of respondents recorded this indicator as relevant to their decisions” (Palmer, 1993:34). In a more recent Australian study using mail questionnaire to senior managers in 100 federal, state and territory agencies, Lee (2008:132) found that within groups, top

management rated costs as being most important; other levels of management perceived outcomes and customer satisfaction as being most important. A preference for cost indicators in public agencies will have far reaching consequences on performance information use. DiIulio (1993), questioning the relevance of efficiency and cost in the public sector, not unlike Waldo's (1984:193) famous "efficiency for what" diatribe, offered the following:

> In many, perhaps most, government agencies, the relationship between valued inputs (people, money) and desired outputs (less crime, better public health) is ambiguous. Where goals are nonoperational, and the technologies necessary to achieve them are either uncertain, or completely unknown, or simply unavailable, the quest for a "bottom line" is a fool's quest. To the extent that goals are vague or inconsistent, the concept of efficiency is irrelevant. The concept of efficiency is thus irrelevant to many, if not most, public management tasks. (DiIulio, 1993:147)

Logically, relying strongly on cost measures, either as a sample of available indicators or because cost indicators are preponderant in the performance indicators in an agency, would contribute to explain why performance measures would be scantly used. Another influence of financial and cost measures on the use of such measures by managers has to do with timeliness. As we have seen before, one of the prime characteristics of performance measurement is its regular influx of data (de Lancer-Julnes, 2006:223). Financial and cost data are linked to the budget, which is a management tool updated on a yearly basis. Therefore, "non-financial measures are more timely than financial ones; the measures are very measurable and precise; the measures are meaningful to the workforce so aiding continual improvement" (Medori and Steeple, 2000:521). For these two reasons, preferences for input (cost) or efficiency data might influence negatively the use of performance measures in managerial activities.

Reporting

Reporting is not only an essential pillar of performance management, it is also one of the functions of the chief executive, as identified in 1937 by Gulick's famous POSDCORB. Reporting is the vehicle through which public accountability is made possible. Public accountability refers to "the reporting of comprehensive information about the condition, performance, activities and progress to all those with social, economic and political interests" (Coy and Dixon, 2004:81). This heightened accountability, as we have seen earlier, would be one of the main benefits of performance measurement (Berman and Wang, 2000:417; De Bruijin, 2002:580-581; Carvalho et al., 2006:174), alongside performance improvement.

Citizens in New York City, when asked in focus groups to describe the types of government reports they would like, said they wanted:

> (1) reports and information presented clearly and simply, (2) honest reports about how government programs are working, (3) all the news, not just good news, (4) to understand the challenges that their government and their community are facing, (5) to know how and where they can obtain additional information about services and key issues, (6) to be able to evaluate information for themselves, without spin, (7) to know what other jurisdictions are doing and how they are doing in comparison (Berman, 2008:6).

Many of these wishes expressed by citizens about government reports have to do with the explanation, justification and contextualization of data. This is in accordance with other definitions of accountability like Rubin's (2005: 2073) definition of accountability as "the ability of one actor to demand an explanation or justification of another actor for its actions and to reward or punish that second actor on the basis of its performance or its explanation."

The common view regarding public agencies relations with the press and reporting is that public agencies are more interested in damage control than attempting to plead their case with their stakeholders and the wider community they serve. One of the motivations behind this defensive stance would be a fear of political fallout. Another motivation is that the release of some indicators, for example indicators on crime, can later negatively impact the performance on some other indicator, like citizen fear of crime (Chapman and Lombard, 2006:793). In a rare study on the press coverage in news stories and editorial of performance measurement report cards of state governments in the United States, Schiffel and Smith (2006) found complex patterns of press coverage of government performance. In terms of the amount of coverage, their results are that "extreme grades and extreme grade changes received the most coverage. The lowest grade received more coverage than good grades, but improved performance received more attention than deteriorating performance" (Schiffel and Smith, 2006:23). In terms of tone, the authors' results are concordant with Berman's (2006:20) mention that the press can and does acknowledge government honesty and forthrightness in reporting, the tone of the articles was mostly positive or neutral; only negative performance brought about a negative tone in coverage (Schiffel and Smith, 2006:23).

From observations in New York City, Berman (2006) found that many government officials were afraid that reporting on how some services levels were substandard would hurt the administration in power. She anecdotally observed that no such ill effect ensued (Berman, 2006:20). An empirical study of 143 cities with over 20,000 inhabitants in Spanish local government gave some support to the uneasiness of government officials.

Brusca and Montesinos (2005) looked at the electoral response of citizens to financial measures of local government. Their findings are worth citing at length:

> In short, the results of the empirical study show that budgetary and financial ratios influence electoral outcomes. The results are more consistent if we use the data from the last year of the term of office than if we use the data from the first year and the change with respect to the previous term of office. Nevertheless, data from the first year of the term office can be relevant too. For example, changes in tax revenues and in other financial charges can affect citizens' votes. (Brusca and Montesinos, 2005:208)

The authors add that "it appears that citizens take account of budgetary management and are not influenced by the results of other policies, such as the level of municipal debt or even the fiscal pressure" (Brusca and Montesinos, 2005:209). It is unclear if the same happens for performance measures in other local services. It is worth noting that political consequences are an innate consequence of managing public organizations: citizens have the right to be informed (*see for example* Wilson (2004:37-38)). This responsibility of disclosure takes a heavier toil for public agencies than private organizations. Not only do public organizations have a legal obligation to report their activities, it is argued that public organizations have the moral obligation to report beyond the minimum required by law (Gordon et al., 2002: 237). It explains why, when Wall and Martin (2003) looked at Irish semi-state organizations, which have a fuzzier status than pure government agencies in Ireland and Northern Ireland, semi-state organizations preferred to act like private organizations in their reporting. Semi-state organizations were found to "offer the least amount of performance information with which the organization could be held to account and to make an informed judgment on how it and its resources were being managed" (Wall and Martin, 2003:507). Still, beside minimum legal disclosure, there is little incentive for managers in public organizations to be fully transparent in their reporting.

The reluctance to disclose performance results, especially results deemed poor, has been captured by Try and Radnor (2007) study of federal managers in Canada. The recurring theme of their interviews of 16 federal executives conducted in 2004, was that "in the face of a highly risk adverse culture, with limited rewards for success but considerable punishment for failure, public disclosure of poor programme results will have limited appeal, which can be predicted to result in selective reporting" (Try and Radnor, 2007:669).

Even though it has been concluded that local government reporting through reports is not playing "the vibrant role of a mechanism of accountability in an active local democracy that was envisaged for it" (Butterworth, Gray and Haslam, 1989:83), it has been hypothesized that citizens would be important users of local government financial reporting even when that reporting is not produced with them in mind (Brusca and Montesinos, 2005:209). A reason why citizens are not the targeted audience in government reports, according to Christensen and Skærbæk's (2007) study of Australian and Danish governments, is partly because citizens "(…) are not involved elsewhere in communication with reporting agencies but, more importantly, because the overflow created by the network result in a resource-spending frame in which principally the central and reporting agencies routinely interact without citizen involvement" (Christensen and Skærbæk, 2007:124-125). Thus, according to the authors, government reports would be written having accountability to other agencies in mind.

Government reports written for other government agencies rather than to citizens are illustrative of the fact that some measures are better suited for external accountability, while others are more useful for internal management.

How the measures for reporting are not the measures for internal management

Transparency and performance improvement are two expected benefits of performance measurement that are often mentioned and included in official policies (for example, Wilson, 2004:37-38). Some are questioning if transparency and performance improvement can be achieved simultaneously (Trosa and Williams, 1996:45). Their argument, is that in many cases, a single set of performance measures could not be used for both objectives. One reason is that transparency/accountability is conceptually opposed to performance/productivity. If we are to trust Halachmi (2005), measurements for accountability and performance improvement would not be compatible with each other. The author put forward the reasons for this incompatibility:

1. Accountability is living up to performance standards that existed when the use of resources/authority was authorized.
2. Accountability is primarily about relationships: Who is superior to whom? Who is answerable to whom? What must be reported and who decides it?
3. Productivity is more than keeping with past trends or marginally improving on them.
4. Productivity relates to progress, innovation, and change, preferably moving to a higher curve rather than moving to a higher point on the same productivity curve.
5. Productivity is about management, adaptation, creativity, and breaking away from the past or from the group, while accountability is about staying within the four corners of the contract.
6. Productivity results from thinking outside the box, while from an accountability point of view, all such activities suggest deviation and a disregard for the rules.

> 7. Productivity involves feeling good about alleged results and having a sense of achievement, whereas accountability is about feeling right, safe, and capable of defending an official (formal) record.
> 8. Productivity has to do with a continuous free-form process of self-examination and an internal search for new insight, whereas accountability involves external scrutiny and a relatively rigid use of pre-established legal or professional standards. (Halachmi, 2005:261-262)

A second reason is that performance improvements happen internally, and that accountability is done externally. Internal and external audiences, but also different internal users within an organization, would not only have different needs in terms of the amount of details they demand (Halachmi, 2005:260; Rosentröm and Kyllönen, 2007:295), but also different levels of sophistication and analytical skills. The case in point for the different needs of internal and external users of performance information has been made by the British *Public Administration Select Committee*, when they reported the results of their national consultation on performance measurement to the House of Commons. In their report, they stated that:

> One crucial point emerges from our evidence; there is an important distinction between performance information used internally, to support management and aid learning, and information put into the public domain to show how well services are performing. The quality of both needs to be high, but what is appropriate for one may not be appropriate for the other (Public Administration Select Committee, 2003:10).

Managers, elected officials and citizens would have widely different needs when it comes to performance information (Wisniewski and Stewart, 2004:224; Smith, 2005:215). One study done by Jansen (2008) is illustrative of the difference in needs and uses of performance information between politicians and managers. Using the case study method for three local social services departments in Utrecht, Eindhoven and Groningen, in the

Netherlands, the author tried to explain the low level of usage of performance measurement. Jansen (2008:179-180) found that in the three cases, the needs of managers and politicians are so different that they received separate performance reports. Even within management, Managing Directors of the municipality's Social Services' Department hardly use the performance report prepared by their department: the reports contain too many details and are tailored to the information needs of the production managers. A managing director expressed his preferences for a performance report "written by the department's production managers, which only focuses on deviations between the department's targets and its actual performance, together with an analysis of these deviations" (Jansen, 2008:180). Similarly, the two politicians interviewed indicated that they would prefer reports that limit process information to processes implying financial or political risks. These politicians made it clear that they are only interested in performance information on processes or outputs, "(…) if this information has serious political or financial consequences, which could receive negative media attention and/or criticism in the media or from the Municipal Council" (Jansen, 2008:180). The result of having performance reports with a single set of precise measures geared internally for performance improvement meant that managing directors and politicians had limited use of the performance reports (Jansen, 2008:183). Managing directors would palliate this situation by collecting their own performance information on an ad hoc basis, as a replacement of the information provided by the performance report (Jansen, 2008:181). All in all, politicians and directors, because of their need to be accountable, wanted to know the extent to which the organization has succeeded in meeting its plans (Jansen, 2008:183).

This made Melkers and Willoughby (2005:188) say that "measures that are included in documentation simply to support reporting protocols may not prove useful or effective for decision making." To that regard, Ammons and Rivenbark (2008:308) went further and stated that according to their observations in North Carolina, performance management initiatives geared toward reporting are unlikely to foster performance improvement.

A possible response to the gap between measures that are internally and externally relevant to an organization, compounded by the cost of collecting data, is trying to develop a generic set of performance measures intended to satisfy different groups (managers, elected officials, citizens) and their differing requirements. The danger of such an approach, according to Wisniewski and Stewart's (2004:224) study of local authorities in Scotland, is that such generic measures will satisfy no one, as the measures will not accommodate anyone's proper needs. A second alternative is to overtly choose a set of performance measures for either a goal of performance improvement or a goal of accountability. However, what can come of it is the unofficial implementation of a parallel measurement system to cover the other function, which brings confusion and resource strain on organizations (Moxham, 2009:753-754). A third alternative is to have a dashboard and collect two sets of data. This alternative to have "one that responds to the public's interest for information and a second that is tightly focused upon performance and managing to outcomes" has been considered by officials in Snohomish County, WA (Stein, 2007:59). A fourth alternative is to collect only one set of data, but to present data in different formats for different users (Smith, 2005:215). However, it is not clear if

pursuing this strategy will not end up facing the same dangers of offering a generic set of measures. The measures would be too general for managers, no matter how they are presented, and not relevant for citizens; or general enough for citizens, therefore not actionable for managers.

Limits of transparency

As we have seen, the two goals, which would be benefits of performance measurement, would be complicated to simultaneously attain in equal measures. It is worth to enumerate here the limits of accountability and transparency to the citizens. It is often taken for granted that once performance measures have been made public in a report, accountability has been achieved. The critical issue here is "do citizens care about what their government does?" More than being a rhetorical question, it has crucial practical repercussions in performance management. After all, if a performance measurement initiative has been put in place, geared toward uses of accountability with more general measures geared for external use, an opportunity cost is barren by the organization. The organization missed an opportunity to put together performance measures that could be used internally by managers. Before we move forward, an argument on the influence of transparency and openness on performance has to be addressed. Proponents of transparency are often quick to offer syllogisms about the virtues of transparency on performance or ethics. While the questions of transparency and ethics are outside the scope of this research, the question of the influence of transparency on performance is not. While the argument that performance would be improved as a result of increased scrutiny might appear convincing at face value, it is difficult to assess it empirically. A

methodologically convincing way to test this hypothesis requires comparative data of agreed-upon measures of performance among units of government. At the macro-level, one such study has been done by Hauner and Kyobe (2008). The authors compile the first large cross-country panel dataset of public sector performance and efficiency. Their sample includes 114 countries on all income levels from 1980 to 2006. All in all, the authors used a sample of about 1,800 country-year observations for education, and about 900 observations for health. Openness was not among the factors that explained achieved performance levels. Hauner and Kyobe's (2008) results are that:

> Openness could be expected to increase performance and efficiency by increasing competitive pressure on the domestic economy, including the government, as well as raising more generally exposure to the outside world, including through skills and technology transfer. We look both at de jure trade liberalization and de facto openness, but none of them turn out significant in the final tested-down specification (Hauner and Kyobe, 2008:17).

More studies would be needed to establish if transparency and openness promote better performance at the micro (e.g. agency) level.

The early view on the degree to which accountability to citizens works, that is, if citizens are aware and use performance information from their government was formulated by Simon and Ridley (1938). The authors felt that citizens might care if the report is tailored for them. To that effect, the authors said that:

> It was once believed that if the administration published each year a complete and encyclopedic account of city activities, governmental efficiency could be appraised by the citizen. This highly optimistic view of the "intelligent citizen" has long been discarded. The voter has neither the time nor the information to interpret undigested statistical tables in an intelligent manner. The reporting official must act as interpreter as well as recorder." (Simon and Ridley, 1938:466-467)

Seventy years later, describing the well-known Baltimore CitiStat performance measurement initiative, Behn (2008) expressed a different opinion. Behn went as far as saying that citizens are not aware of Baltimore's performance strategy: they care about better services, not how these services materialize. In his own words, Behn (2008:36) said that: "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." The lack of interest and use of information on the part of citizens flies in the face of Berman's (2006:18) assertion that citizens want information about their government. Studying the actual use of electronically available performance information data from Dutch public schools and hospitals, Meijer (2007) found that citizens do not use them. The authors' findings went as followed:

> Additionally, the 'consumers' exhibit little interest in the information on school and hospital performance. In both sectors, the information on the Internet was accessed by citizens but hardly used. Citizens made little use of the information in their choice for schools and hospitals and even less use of the information to ask questions about school performance. The information on Web sites enables accountability to citizens but the opportunities are not used much. One could argue that citizens will need to go through a learning process and that, in the near future, they will start using the information more. However, experiences in countries with a longer tradition – mainly Anglo-Saxon countries – seems to counteract this argument (Meijer, 2007:180).

Meijer's study of how Dutch citizens do not use available performance data to choose schools or hospitals might simply reflect a cultural characteristic. One might recall Brusca and Montesinos' (2005:208) findings that citizens in Spain let financial performance of local government influence their voting behaviors. Opposite findings happened in the world's most sophisticated municipal benchmarking system in England, where "naming and shaming underperforming authorities undoubtedly grabs the attention

of senior officers and councillors, but there is no evidence that local people have taken much notice of CPA scores or punish those leading poor performing councils at the ballot box" (Downe *et al.*, 2007:8). Cultural differences might explain the extent to which citizens use performance information, if at all. For the purpose of this study, it is interesting to acknowledge citizens' appetite for performance information in a North American context.

One of the rare studies of actual use of performance information by citizens in North America has been done in Pennsylvania. Like Meijer's (2007) study, Schneider and Epstein's (1998) study focused on health care: more specifically on the influence of performance measurement in the choice of a cardiac surgeon in the state of Pennsylvania. The main difference between the two studies is the nature of the sample in both countries: public healthcare in the Netherlands, and private health care in the United States. Schneider and Epstein (1998) analyzed the use of information by patients who underwent coronary artery bypass graft (CABG) surgery. Given the dramatic nature of CABG, one would expect that a patient planning such a surgery would be particularly motivated to use available information in the *Consumer Guide to Coronary Artery Bypass Graft Surgery*: a risk-adjusted guide that received extensive media coverage. Schneider and Epstein (1998:1638) described this guide as being a widely distributed guide that was sent to hospitals, surgeons, public libraries, business groups, and the media. Individuals could, at the time of their study, freely order one of the 15,000 available copies. Schneider and Epstein (1998) interviewed 474 Pennsylvanians who underwent a CABG surgery. The results of their study are similar to Meijer's: service users did not use

available performance information to make their choices. What is really surprising in Schneider and Epstein's (1998) findings is the extent to which available information, even by users with direct incentive to be informed, goes unnoticed.

> Ninety-three of the patients (20%) were aware of the Consumer Guide, and 56 (12%) of those said they knew of it prior to their operation. Two thirds of these patients (n=37) had only heard of the guide, while one third (n=19) had actually seen a copy. Eighteen (4%) reported knowing the hospital's categorical mortality rating (higher than, lower than, or within the expected number of deaths). Eleven(2%) reported that the information influenced the choice of hospital, but only 4 of these knew the correct categorical rating, which amounted to less than 1% of all respondents. Only 6 (1%) reported discussing the ratings with a physician. (Schneider and Epstein, 1998:1639)

Contrary to many public services that are paid for indirectly through taxes and dispensed to a limited portion of the population, cardiac surgeries concern one individual, and in the U.S., are paid for through his/her insurances. Only one percent of post-surgery patients could accurately recall performance information for such a service.

Transparency in the public sector is important. The case for transparency for government has been eloquently made at many occasions (for example, Piotrowski and Rosenbloom, 2002; Piotrowski, 2003; Rosenbloom and Piotrowski, 2005). The point here is that as far as performance measurement is concerned, managers faced with a choice between measures for accountability and measures for internal performance improvements should be aware of the shortcomings of accountability, especially accountability to citizens. Measuring and publically releasing that data, is arguably easier than collecting data that can eventually be used to diagnose problematic operations that can possibly later be improved at the margin of diminishing return. However, as we have seen, the accountability benefits of performance measurement in the public sector might not

materialize because citizens, by large, would not care and would not use performance information to guide decisions. In that scenario, not only would an agency not be in effect more accountable, but it would have missed an opportunity to have performance data tailored to foster performance improvement.

Complexity

Managing the public sector is complex. Public agencies have a tall order to fulfill, they are expected to: (1) effectively tackle and curb problems that are not well understood; (2) do so in an efficient manner even though private sector organizations passed the opportunity because there is no profit to be made; (3) operate in an equitable manner that also *appears* to be equitable to all groups; (4) respect due process; and (5) deal with media that focus on government failures and frequently fail to celebrate genuine successes (Holzer and Charbonneau, 2008:9). To really have performance measurement encompass the complexity of government to a large number of stakeholders, the number of performance measures would be high to reflect this complexity (Radnor, 2008b:317).

As we have seen, performance measures would need to be fine grained for managers to be able to act upon them. However, when it comes to measuring the performance of complex operations in the public sector, there is considerable pressure on public organizations to use a few simple measures that can be easily reported to politicians and citizens.

> The increasing econometric and statistical complexity seems to have become a problem for actors that are not part of a profession or dealing with measures on a more or less daily basis. Typically, these actors are citizens and politicians. These

> users "outside the box" no longer understand the technology behind the results (Van Dooren, 2008:20)[†].

A common way to make performance information accessible for reporting to lay people, politicians and citizens, (1) assess performance in a holistic manner, and/or (2) report data with easily categorical interpretations of what the results mean. The latter often takes the form of arrows, traffic lights and smiling faces; the former results in aggregating multiple measures into indices. Again, this strain of performance measures designed for external accountability makes it difficult, if not impossible for managers to act on this information to improve performance. To that regard, Bird and colleagues (2005:19) offered that "(…) complexity or difficulty of public understanding should never be an excuse for insufficiency of analysis". As we will see later on, this includes having performance measures that take into account the socio-economic characteristics that are outside of public managers' control. This lack of acknowledgement of complexity can ultimately doom a performance measurement initiative, as managers will not use simplistic performance measures. This is one of the reasons for the demise of the defunct *Law Enforcement Assistance Administration*'s *Standards and Goals Program* that went on in the 1970's in the American criminal justice system (Zedlewski, 1979:490).

The negation of recognizing complexity in performance measures by amalgamating performance measures is problematic. The most extreme case is the report card-type assessment of an entire municipal government with a single measure, often a letter grade. For all its intricacy of the evaluation scheme, the early versions of the *Comprehensive Performance Assessment* (CPA) system in Britain ultimately boiled down the measures

[†] Van Dooren (2008:20) was quick to add that: "Statistical sophistication sometimes puts up a smoke screen for inadequate conceptualizations, poor quality data or hidden agendas"

of a municipality on a single measure. Reporting "single snapshot judgments of the performance of the whole authority, on the apparent assumption that that performance will be uniform across its dozens of service areas and thousands of employees" (Game, 2006:472), went against the spirit of the measurement before the index: the painstaking performance measurement of the many services, the inspections of the Best Value system, the auditing. Coming up with a single amalgamated score for the performance of a municipality rests on the assumption that the services are correlated among themselves. Studying local authorities in England, Boyne (1997) offered this remark:

> There is very little evidence to suggest that levels of performance vary together across services. For example, an authority which processes benefit payments quickly is as likely to deal slowly as rapidly with planning applications; and there is no tendency for councils with a high-quality housing service also to have a high-quality refuse collection or planning service. These results imply that performance is not driven by the general characteristics of local councils, but by the circumstances, organization or ethos of specific service departments. It is therefore inappropriate to categorize councils into 'high performing' and 'low performing' groups across all services (Boyne, 1997:40).

Beside this assumption of correlated performance in services, an overall performance assessment, especially if it is naïvely weighted (see the upcoming *weights* section), means that a very poor performance in one service can be balanced by one or more highly performing services.

More baffling than the CPA example is the evaluation of performance by a single measure, without relying on the amalgamation of different criteria. This happens often in human resources management, where to ease the burden of the rater, a pass fail system in put in place. Such a performance appraisal systems lacks the capacity to measure fine distinctions in employee performance (Roberts, 1996:366).

Weights

To fulfill goals of reporting to non-specialists, performance measures are sometimes summed up in indices. The straightforwardness of the amalgamating process takes away, at times, from the complexity of what is being measured. Nowhere is this more patent than in the weighting process. The tendency to naïvely weight measures was recognized early on in the performance measurement literature, with Ridley's (1927:36) remark on weight in policing performance measures: "Simply to add and compare the totals of complaints and arrests is unreliable; and yet this is often done. Instead, each class of crime should be weighted on the basis of its gravity or seriousness." By naïvely weighted measures, we mean the explicit or implicit practice of using the arithmetic average of all components of an index. That is to say that a naïve index values on an equal footing all its parts. This is summed up by Straight (2000):

> In too many situations when using multiple measures, weights are not considered with the result that all measures are given equal weights. Such systems may not produce the results that managers desire and expect since it is unlikely that there will be true equality between several measures in any metric system. Equal weighting sends the wrong signals to the workforce when certainly some measures are more important than others. (Straight, 2000:512)

The choice of a weighting system, or more possibly the lack of reflection on it, can have a significant impact on the value taken by the index (Stone and Davis, 2007:217-218). This in turn can send a distorted image of an organization's performance. The torts being done to an organization can be that managers will (1) recognize the naïve nature of a composite performance measure and ignore it, or (2) take actions on an index that does not truly reflect the performance of an organization. To palliate to this problem, Jacobs and Goddard (2007:109) suggested that in the face of an ad hoc, arbitrary choice of weight, "greater attention should be paid to the origin and nature of weights and the

sensitivity of composites to changes in the weighting structure." The methodologically proper directions of how to weight an index is outside the scope of this dissertation[‡]. All in all, in this subsection, an argument has been offered that complexity might be the first victim of crudely fabricated measures for reporting purposes to non-experts like politicians and citizens. As we mentioned above, aside from naïvely weighted indices, another tendency for publically constructed performance measures is to go with arrows and traffic lights.

Arrows and traffic lights

We explained earlier how assumption of correlated performance in services shapes a simple performance index. Specific services are also measured by a simplistic overall performance measure. One example of this phenomenon is the measurement of overall performance of police services in Swedish municipalities with a color code. Color coded performance measures might ease reporting to lay people, but it provides little help for managers to foster performance improvement.

> Measurement of policing through colours, we argue, poses insurmountable obstacles to capture the richness and variety of police work. Therefore, as aggregation progresses, most dimensions turn into yellow, and "yellowness" diminishes the potential of the scorecard for purposes of operational improvements at the local level. (Carmona and Grönlund, 2003:1492)

In that same study of municipal police services in Sweden, the authors interviewed police chiefs. What came out of it was an annoyance from the police chiefs toward the use of simple color codes as performance measures: it is seen as very impractical to assess the situation in their districts. Swedish police chiefs confided to the authors that they found

[‡] Readers wanting a detailed step-by step protocol on how to weight an index should consult Alden (2006:8-9).

problems assessing the performance of individual districts: “If a district reports red on two scorecards and green on one and yellow on the others, what is then the overall picture of the district?” (Carmona and Grönlund, 2003:1487).

Additionally, the authors observed a recurring theme in their interviews of police chiefs.

> Many pointed out that data from the scorecard should be complemented with information produced by the central information system (such as the crime statistics) and accounting reports (e.g. compliance with budgeted expenditures), and this would produce a fair picture of a district’s performance. (Carmona and Grönlund, 2003:1487-1488)

The use of crude measures like traffic lights and smiling faces is not limited to Sweden. ‘Happy smiling faces’ and ‘sad faces’ have been used to measure improving and deteriorating performance in municipal public housing in Wales (Davies, 2004:37). Academics are sometimes the ones calling or supporting such measures, like Voyer (1999:263-264) in the case of Quebec.

In all fairness, it would be biased to bestow the sole responsibility of the neglecting of complexity on citizens and politicians, for whom managers would have to simplify greatly the measures by which performance will be judged. There are reported cases where managers were unwilling to reach further than most easily accessible data. For example, from interviews of fiscal/budget officers from state agencies in Georgia, Lu (2008:17) remarked that, “if a measure needs a substantial amount of administrative work in collecting data, then, as one fiscal officer straightforwardly noted, ‘We will not do it’.”

Disaggregated data

The need for disaggregated data in public performance management is discussed in the literature as another item needed to capture the complexity of public services. In the early

years of the 1990s, the over-aggregation of performance data was already identified as one of two major impediments of performance information management use, along with timeliness problems of data (Wholey and Hatry, 1992:609). Wholey and Hatry (1992:609) stated on that topic that "managers need information on outcomes for their own area of responsibility and for various client groups, not simply for the program as a whole." Since government agencies are entrusted with pursuing the public good and pursuing social equity, the accessibility of disaggregated data becomes a necessary condition for performance measurement in the public sector. First, it offers staff a chance to learn more if data is in its "(…) original, most unfiltered, disaggregated form" (Behn, 2008:15). Second, having performance measures broken down into groups makes it possible to observe if there are gaps between the groups. This need to have disaggregated performance information on more traditional effectiveness of efficiency indicators, to inform managers about disparity between groups, has long been recognized in areas such as public education (Scollay and Everson, 1985:207; Wilson, 2004:44) and public health (Smith, 2005:214-215).

Being able to compare citizen satisfaction among groups has also been deemed as an important step that needs to be taken by public organizations in the literature (Kelly and Swindell, 2002:272). Liederbach and colleagues (2008) summarized it best:

> The negative assessments of traditionally disenfranchised groups – even within the context of positive overall ratings – need to be recognized as a significant concern. Dissatisfaction among a comparatively small proportion of residents can work to block programmatic success, especially when that dissatisfaction emanates from within neighborhoods that have been the target of community-oriented reforms because they have historically been antagonistic to the police. (Liederbach *et al.*, 2008:285)

The access to disaggregated data is not a sufficient condition for the measurement of social equity; it is merely a necessary condition. Disaggregated data have direct implications on target setting and service delivery. The English Public Administration Select Committee, when reporting to the House of Commons, made it clear that:

> A national target can be met in more than one way, and some of them promote greater equity while others do not. For example, a 10% improvement in services can be achieved if all providers improve equally. Alternatively it can also be achieved if some units do disproportionately well while others fail. If top performers improve most, this will widen the gap between citizens in different parts of the country, while if poor performing agencies do best, this will not only raise the average but also reduce inequalities. It is important therefore to be clear about objectives (Public Administration Select Committee, 2003:13).

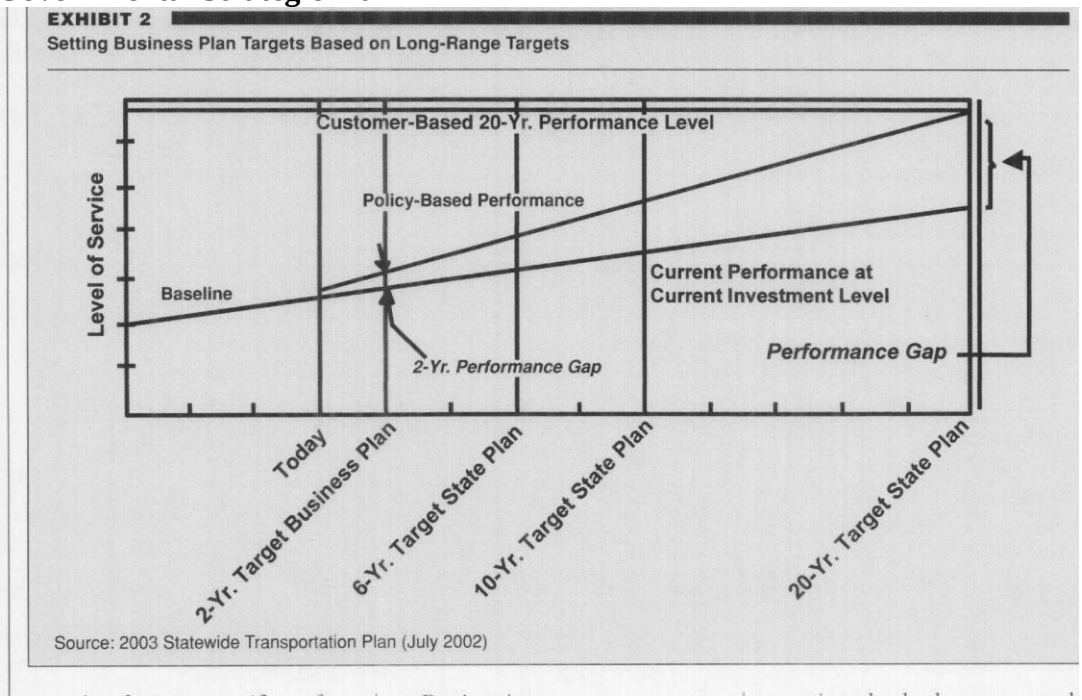
This is to say that an absence of complexity in the aggregation of data can have serious effects on performance and social equity. Overly aggregated data, along with naïvely weighted indices and simplistic performance measures like traffic light arrows and smiling faces rob performance measures of their representativeness of complex public services. In the next section, we will examine what the literature has to say about the very nature of performance.

Non-linearity

Despite many warnings from the literature, the very nature and shape of performance is often overlooked in performance measurement initiatives. The result is that complexity is disregarded. The former Minnesota's *Statewide Transportation Plan*, which was completed in 2003, is a patent example of the implicit performance linearity assumption that goes on in many public sector performance measurement initiatives. In its ten-point policy, the Department of Transportation (DoT) of Minnesota spells out on its point 8, that in order to make the Mn/DoT better, it seeks to "continually improve Mn/DoT's

Internal Mangement and Program Delivery" (Mn/DoT, 2003). More telling is the figure that was contained in that plan. Figure 1 was originally in the 2003 plan, which is no longer available in its complete form; it has been reproduced by the head of the performance measurement section at the Mn/DoT in a 2003 article.

Figure 1. Example of the Implicit Performance Linearity Assumption in a Governmental Strategic Plan



Source: Feit (2003:43)

According to the 2003 plan from MN/DoT[§], constant performance improvement is possible. Given the assumed shape of performance, performance improvement is thought to be linear. Linearity, in a performance setting, means that no matter the performance at a given time t_x, it is possible to increase performance at t_{x+1} at an equal rate than between t_{x+1} and t_{x+2}. Moreover, this means that the same effort, no matter the performance

[§] To be fair with the Minnesota Department of Transportation, their new *Minnesota Statewide Transportation Policy Plan 2009-2028* does not contain that kind of linear assumption about performance. For example, Figure 4.25 Minnesota Annual Vehicle Miles of Travel (VMT) (Mn/DoT, 2009:4-28) includes both linear and logarithmic predictions.

baseline at t_x, will yield the same gains in performance. As we will see later, this assumption has important consequences for the setting of targets and for sensemaking derived from expected and observed performance.

By design, the assumption of linearity is difficult to correct in a single agency performance measurement and improvement initiative: there is little information to determine the relative performance of the agency when it lacks external data. However, even in complex municipal benchmarking systems such as the 'Best Value' component of the former English CPA system, diminishing returns in performance improvement can be ignored and constant improvement can be expected (*see* Bowerman, Ball and Francis, 2001:324). While studying that system, Andrews and colleagues (2005:652) found that inspectors were more likely to expect a possible performance improvement in already high performing agencies.

This assumption that performance improvements can be attained in constant strides becomes even less realistic when multiple agencies are evaluated with a given set of criteria. In terms of performance, where agencies stand, matters. It is very different to improve performance in a poorly performing agency than to make an excellent agency even better: "Organizations at the bottom end of the distribution have a lot of room for improvement, whereas those at the top end may need disproportionate managerial skill or effort to perform even better" (Boyne and Chen, 2007:461). Commenting on the state of the literature, Jas and Skelcher (2005:208) remarked that "although this point appears

obvious, the literature on performance improvement in the public sector is written around a normative model that takes little account of this observation."

In the present review of the literature, we found some instances where the non-linear shape of performance was acknowledged. For example, the impracticality to infinitely increase performance has been expressed or identified for municipal government (Higgins, 2005:459-460; Boyne and Chen, 2007:468,473; Behn, 2008:38), policing (Loveday, 2006:286), fire services (Wallace, 1977:31), passports and visas processing (Jones, 2001:494), public hospitals finances (Alexander, Weiner and Griffith, 2006:1023) and public health (Castelli *et al.*, 2007:114). Jones' (2001) statement on the effect of the linearity assumption on improvement targets summarizes the essence of the idea expressed by these researchers:

> Although there may be a presumption that performance targets should be improved, in many cases further improvement is difficult if not impossible. Indeed, where significant advances have already been made, it becomes increasingly more difficult to continue to accomplish ever higher targets, and there is a natural point where higher targets can no longer be set, especially within existing resource availability. (Jones, 2001:494)

This phenomenon known in economics as "decreasing marginal rate" is present in any undergraduate *introduction to micro-economics* textbook. Similarly to other simplistic assumptions about performance reviewed earlier in this research, the linearity assumption can have important consequences on administrative operations. Boyne and colleagues (2006:304-305) explained that "an invalid assumption of linearity could result in inappropriate practical results. After all, a relationship that is subject to diminishing returns could produce an overly optimistic policy prescription if the diminishing returns are not incorporated into the analysis."

Conscious about overpromising systems and unrealistic performance goals, two researchers from Idaho National Laboratories tried to explain how performance actually improves over time. Harbour and Marple (2005) tried to answer this basic question by plotting around 1,200 data sets of individual performance records from diverse settings, such as records from sports event, global public health care and offshore oil discovery. Their overall findings are worth presenting here:

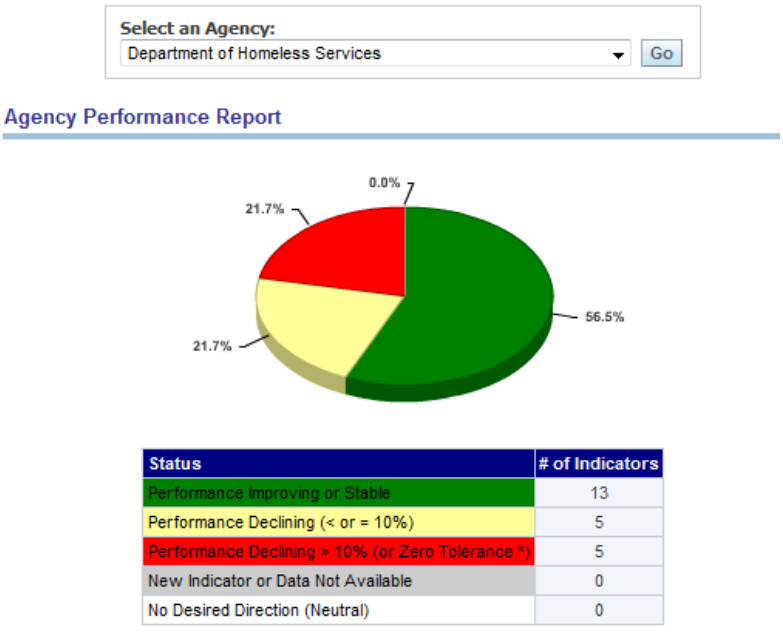
> In summary, whether we're talking about human health, human speed, oil discovery rates, or powerboat speed, mature performance curves seem to follow the same characteristic pattern, albeit at highly varying magnitudes. An initially steep relative rise in performance gain is followed by a pronounced slowdown. Such resultant slowdowns, characterized by ever-smaller gains at correspondingly longer intervals, create a terraced, step-like, or flat-top appearance in curve form. (Harbour and Marple, 2005:16)

These findings about the non-linearity of performance corroborate the observations of public administration researchers in fields of municipal government, policing, fire services, passports and visas processing, public hospitals finances and public health presented above. Unlike Harbour and Marple's (2005) results, few public administration researchers comment on the actual shape of performance, albeit they might mention that it is not linear and that there are natural thresholds to performance improvement (Bititci, Turner and Begemann, 2000:697-698). Notable exceptions in the literature are the explicit curvilinear relationship between the contracting out capacity and performance improvement in Taiwan (Yang, Hsieh and Li, 2009:690), the citizen satisfaction with municipal services in New York City (Van Ryzin and Charbonneau, forthcoming), and room wait-time in English National Health Service hospitals (Kelman and Friedman, 2009:928). The case of waiting time in English National Health Service hospitals is

especially explicit, as the authors recognized that increasing the amount of patients who wait less than two hours from 76% to 88% can be an equivalent accomplishment to improving waiting times from 96% to 98% (Kelman and Friedman, 2009:928).

Given the widespread presence, in performance measurement initiatives, of more obvious simplistic devices like naïve weights, arrows, traffic lights and smiling faces than aggregated data, it should not be surprising that few researchers and fewer practitioners take into account diminishing return. Even when some accommodations are made for the curvilinear shape of performance, like New York City's *City Performance Reporting* (CPR) performance measurement system, the formula is not fitted to specific services. New York City CPR is an example that is particularly interesting for this research. It has a built-in sensemaking feature for its measures, in the form of traffic light color code.

Figure 2. Example of the Implicit Performance Non-Linearity Assumption in a Performance Measurement Initiative – NYC CPR



Source: NYC CPR (2009)

The CPR system tracks variations in performance. It allows some room for non-linearity, as it considers that *performance improving or stable* is *green*. Furthermore, performance decreases can be either *yellow* or *red*. Complexity is somehow denied, as all 44 departments and agencies in New York City are evaluated with the same cookie cutter slide rule. Quite arbitrarily, a negative performance variation of less than 10% is deemed yellow and more than 10% is deemed red. This example is an illustration that the uses of simplistic devices in performance measurement initiatives are not correlated. Some aspects of complexity can be taken into account while others go unnoticed.

Performance measurement is an information-based tool that tries to simplify a complex reality into a manageable number of graspable measures. The argument in the literature is that the level of complexity needed by managers for internal performance improvement is different than the level of complexity suited for reporting to politicians and the public. In this section, we reviewed what the literature perceived as excessive simplification of reality. In the next section, we will see what the literature has to say about the shortcomings of performance measurement.

Shortcomings of performance measurement

Performance measurement, like any other management tool, has shortcomings. The same precautions disclaimed for the alleged benefits of performance measurement apply for the alleged shortcomings. Depending on one's general biases for or against the use of a positivist and rational tool to inform and possibly guide decisions, the shortcomings identified by deductive or inductive research will appear more or less convincing.

Opponents to performance measurement might be more receptive to findings of performance measurement shortcomings from essays, normative and grey literature than proponents. Cutting through the ontological Gordian knot is beyond the scope of this research. We will present here the perceived shortcomings of performance measurement from a plethora of sources. One of the better known malfunctions of performance measurement, namely gaming, will also be discussed.

It is challenging to present shortcomings of performance measurement without degenerating into simply stating numerous lists of recriminations. Recriminations do not often come in single units. To better grasp the shortcomings of performance measurement, Halachmi's (2005:264) identification of the reasons why performance measurement initiatives may not be successful, is useful because it permits a classification of the perceived faults inherent to performance measurement. First are reasons related to human behavior, like the alleged blocking of ambition and innovation (De Bruijin, 2002:581-583) and loss of satisfaction on the part of employees as "(…) freedom and expertise have been taken from them, and a routine has been imposed instead" (Hoogenboezem and Hoogenboezem, 2005:575). Second, some reasons are related to the nature of governmental organizations, like the so-called process-burden and overmanagement that can come from performance measurement (Purdue, 2005:124). Third, other reasons have to do with the underlying theoretical bases of performance measurement like "the shaky assumption at the heart of performance measurement that studying the past is a sure way to navigate into a better future" (Halachmi, 2005:264). Another example of theoretical grieves are tendencies of reification (Van Peursem, Pratt

and Lawrence, 1995:60; Bovaird, 1996: 162). Lists of performance measurement shortcomings can include elements from one or more of these three categories.

The most dramatic depictions of the shortcomings of performance measurement come from narratives. They consist of comparisons of limitations to a mythical monster (Radin, 1998) and to diseases (Bouckaert and Balk, 1991). The names of the identified diseases of performance measurement are quite telling: (1) Pangloss disease [if it wasn't efficient, wouldn't we have changed it already], (2) impossibility disease, (3) hypochondria, (4) convex/concave disease, (5) hypertrophy, (6) atrophy, (7) Mandelbrot disease [if we put more police on the streets, you will measure more crimes], (8) pollution disease, (9) inflation disease, (10) enlightenment/top-bottom disease, (11) time-shortening disease, (12) mirage disease, and (13) shifting disease (Bouckaert and Balk, 1991:230).

A seminal list of performance measurement shortcomings originates from an essay by Perrin (1998). The author identified eight limitations of performance measurement. The list goes as follows: (1) varying interpretations of the "same" terms and concepts, (2) goal displacement, (3) use of meaningless and irrelevant measures, (4) confusing cost savings and cost shifting, (5) obscuring critical subgroup differences disguised in aggregate indicators, (6) limitations of objective-based approaches to evaluation, (7) uselessness for decision making and resource allocation, (8) reduced focus on outcome as a result of narrow-focused management objectives (Perrin, 1998:370-375). Anything short of a total reproduction of Perrin’s explanation on each of these perceived limitations will not render justice to the author’s argument: therefore, we suggest a systematic rebuttal of Perrin’s list that was shortly after offered by Bernstein (1999). According to this author,

the shortcomings listed by Perrin are not innate to performance measurement, but rather describe faulty implementation from "poorly implemented systems that focused too much on process and 'collection for collection's sake', as opposed to appropriate use of appropriate measures" (Bernstein, 1999:86).

Determining the validity of perceived shortcomings of any management tool is arduous. This is especially so when management tools or concepts have a virtuous component in their *raison d'être*. Performance measurement fosters rationality; citizen participation furthers democracy; representative bureaucracy embodies equality. When management concepts are defined in a virtuous way, they can become impervious to critiques in their proponents eyes. If any dysfunctions associated with their use or practice are pointed out by opponents, proponents can simply dismiss these critiques by stating that the implementation of their tool or concept is faulty. Such concepts become impossible to falsify.

Gaming

Gaming is the managerial practice of excelling on what is being measured at the detriment of the activities and services that are being measured. Smith (1995:298) defined gaming as the "deliberate manipulation of behavior to secure strategic advantage. Thus, while misrepresentation leads to distortions in reported behavior, gaming is the equivalent manipulation of actual behavior." There are different degrees of gaming. It can range from inoffensive measures like developing margins or buffers between standards and targets to reduce the risks of failing to meet requirements (Heinrich, 1999:357), to

outward lying and actively falsifying operations like having police officers solve fictitious crimes to bolster crime statistics (Young, 1991: 324, 326-327).

Gaming, like other perverse organizational behavior as corruption or plagiarism, is an inherently difficult topic to study. Empirical evidences are likely to come from in depth case studies. By design, gaming is difficult to detect. It can pass as poor or great management, poor or great performance. Sometimes, there are few perverse effects. For example, one way to game performance measures in the field of policing "consists of focusing efforts on crimes like drug use that is by definition identified and cleared simultaneously" (Roche, 2008:337). Survey research is ill-suited to measure it. These methodological limitations can probably explain the lack of systematic empirical studies on this topic. Nonetheless, this does not make gaming any less real.

On theoretical grounds, Radnor (2008b) offers a typology of four kinds of gaming: muddled, massaging, manoeuvring, or manipulated organizational gaming. By "muddled", Radnor (2008b:324) refers to organizational gaming that occurs (1) due to the lack of a review process, or (2) "when poorly defined performance indicators (PIs) mean that it is possible to creatively interpret or spin the output reported metric." Organizations are asked to collect and report lots of measures that are not used or are meaningless. By "massaging" organizational gaming, the author describes situations where "the target or PI is understood but the base data is either not available, not recorded or not reported in order to hit the target i.e. the data is massaged or manipulated" (Radnor, 2008b:324). "Manoeuvring" organizational gaming happens when performance indicators or targets are "clear and understood but the activities are

(creatively) developed or implemented in order to achieve them? (Radnor, 2008b: 324-325). Finally, "manipulated" gaming is when as a result of focusing on the indicators or targets, activities are deliberately changed, implemented and even encouraged in order to "hit the target" even if it results in poor service delivery and outcomes for the actual recipients. (Radnor, 2008b: 325).

Radnor's (2008b) list of outcomes is not without calling Bevan and Hood (2006) four outcomes of performance management with targets. According to these authors, pressuring organizations with performance management targets can result in four scenarios with different levels of gaming. The four scenarios are:

1. All is well; performance has been exactly as desired in all domains (whether measured or not);
2. The organisation's performance has been as desired where performance was measured but at the expense of unacceptably poor performance in the domains where performance was not measured;
3. Although reported performance against targets seems to be fine, actions have been at variance with the substantive goals behind those targets (hitting the target and missing the point);
4. Targets have not been met, but this has been concealed by ambiguity in the way data are reported or outright fabrication. (Bevan and Hood, 2006: 420-421)

All in all, gaming might be the most serious shortcoming resulting from performance measurement. It is difficult to detect and difficult to assess how much of it is going on.

One often overlooked shortcoming of performance measurement has to do with the context in which it operates. As we have seen, one of the main aims of performance

measurement is improvement of performance, usually effectiveness or efficiency. The constant financial crisis in which many government agencies operate might explain some critiques' point about the modest improvement from performance measurement initiatives. Halachmi (2002b:234) summarized it best: "(…) given the constant reduction in the amount of resources available to public organizations in recent years, in comparison to the relative increase in their service load, most government agencies have already been left with very little slack." This ties in with the topics that were covered earlier in this review of literature on performance measurement. Given, as we have seen, that the shape of performance itself is often logarithmic rather than linear, and that because of repeated cuts and constant fiscal stress, public agencies are lean, this would mean that possible performance improvement is marginal. We mean here marginal in its mathematical signification and not how it is referred to in the vernacular. If performance measurement indeed brings about gains in term of efficiency (but also of effectiveness) for already efficient agencies, the gains would be remarkable. However, they would be small if measured in simple arithmetic variations from previous periods. Looked at from simplistically designed performance measurement frameworks, the gains would be seen as disappointing. This could translate into questioning the very use of performance measurement. What could help providing more precise data to depict a public agency would be comparable data from other public agencies. This is precisely what we will cover in the next section.

Benchmarking

Basics of benchmarking

Benchmarking is a branch of performance measurement. Etymologically, benchmark was coined in 1842 to describe the practice of marking the exact emplacement on the ground of a bench supporting an observer's instruments, so that subsequent measurement would be done from a strictly comparable basis (Moriarty and Smallman, 2009:486). The ability to establish best, or at least better, practices through external comparisons is one of the strengths of benchmarking (Raaum, 2007:43). Benchmarking falls into what Heinrich (1999:367) described as useful performance management systems, which "(…) improve programs by assisting managers to identify poor performers, to follow up with corrective actions, and to reward good performers and replicate their approaches."

Modern benchmarking in the public sector is believed to have emerged in the 1970s in the U.K. through the development of comparative studies between authorities, performed by local authorities themselves (Bowerman, Ball and Francis, 2001:324). In the private sector, the Xerox Corporation is often credited as being the first organization to systematically use benchmarking in its operations around 1979 (Kouzmin *et al.*, 199:123; Bowerman *et al.*, 2002:432). A definition of benchmarking in the private sector was offered by Camp (1989a, 1989b). Benchmarking would be the positive, proactive, continuous process of measuring against the best to change operations in a structured fashion to achieve superior performance (Camp, 1989a:62, 1989b:75). Public agencies, like *Her Majesty's Treasury*, acknowledged that benchmarking has been useful for public and private organizations alike (HM Treasury, 2003:33).

Notwithstanding, important differences would exist between benchmarking in the private and public sectors. The ethos of what constitutes cooperation and competition in both sectors is at the heart of it. When studying NHS public hospitals and clinics in the U.K., Guven-Uslu and Conrad (2008) summed up the differences in essence.

> Particular issues are the contrast between the confidentiality ethos surrounding benchmarking in the private sector and the requirement for openness in the public sector. Another contrast is the 'no-blame' culture surrounding benchmarking in the private sector—the objective is continuous improvement. In the public sector, by contrast, the publication of league tables is being used to 'name and shame' poor performers (…) (Guven-Uslu and Conrad, 2008:240-241)

Differences in issues of confidentiality, accountability and transparency between the private and public sectors change the dynamics of how the information is perceived and used. Learning from others is an integral and sometimes hard part of benchmarking. Sharing information implies cooperation; rankings, which are often but not always a part of benchmarking (Holloway, Francis and Hinton, 1999:353), imply competition. This makes Braadbaartand and Yusnandarshah (2008:423-424) say that "benchmarking is, in this perspective, a zero-sum game, where one organization can stay ahead of its peers by guarding rather than sharing trade secrets." To that effect, Bowerman and colleagues (2002) presented an insightful reflection on this confidential/transparency duality. Despite its size, it is worth citing at length.

> Information generated through benchmarking in the private sector is confidential to the organisation. The corollary of benchmarking in the private sector being internal and voluntary is that benchmarking data are confidential to the management of the organisation and it is usually considered important that they should not be disclosed outside the company. Indeed, maintaining confidentiality can prove to be a major difficulty in establishing benchmarking relationships (…). For this reason private sector organisations sometimes subscribe to benchmarking "clubs" which both maintain anonymity and facilitate the exchange of commercially sensitive data. (Bowerman *et al.*, 2002:434)

The authors continue their reflection for the public sector.

> By contrast, in the public sector notions of confidentiality are antagonistic to accountability. The generation of comparative data as part of the process of benchmarking is inherently linked to issues of accountability; if performance data is available for a public body, should not the taxpayers, or local constituency, have a right to know? Such "rights" would tend to justify compulsion. (Bowerman *et al.*, 2002:434)

In summary, benchmarking in the private sector is achieved when firms cooperate with a selected few by sharing operational data with each other. The product is the availability of comparative data that will remain hidden to firms outside of the benchmarking club. The result for the firms in the benchmarking club is data identifying operational strengths and weaknesses that can give a competitive edge for the benchmarking firms. Learning from the data happens away from the public eye. All in all, cooperation in benchmarking fuels future competition.

Benchmarking in the public sector can happen on a voluntary or on a mandatory basis. Benchmarking tries to introduce competition into a "state apparatus context that is characterized by the cooperation of public sector agencies for the 'collective' public good" (Kouzmin *et al.*, 1999:125). The result for the public agency in a benchmarking system is data identifying operational strengths and weaknesses with varying levels of transparency, depending on who collects the data. If the benchmarking system is voluntary, public agencies might follow the private sector's example of secrecy and keep the results of the benchmarking effort limited to the participants. If the collecting organization is a governmental organization, like the national government or a state/provincial/territorial governmental agency, the level of transparency can also vary.

Some government bodies are secretive and do not share data outside participants, for example the ministry responsible for municipal affairs in the Canadian provinces of Ontario and Quebec. Other government agencies embrace transparency to the citizenry, for example the ministry of municipal affairs in the Canadian province of Nova Scotia, the Audit Committee in the U.K., the National government of Norway, the territorial government of New South Wales in Australia.

According to the literature, following the private sector recipe for benchmarking in the public sector does not work. First of all, finding partners for benchmarking is challenging, as public organizations pursue very different goals from one another. It made sense for Xerox to compare its operations to firms like L.L. Beans that are outside the copier business: they both pursue the maximization of their profits and their market shares. It makes less sense for a public health agency to compare itself to a municipal government, as they serve very different tasks, use different measures and have different goals (Kouzmin *et al.*, 1999:125). It might also be difficult to find partners in the public sector when participation to a benchmarking club is voluntary. Examples from long running benchmarking initiatives in the United States, like the North Carolina Benchmarking Project and ICMA's performance measurement initiative, illustrate this difficulty. At present, the North Carolina Benchmarking Project includes 17 out of the 551 (or about 3%) municipalities in the state. Likewise, ICMA's voluntary systems' membership includes 172 American villages, boroughs, townships and cities out of 35,933 general-purpose local governments (or less than .5%); and 25 out of 3,034 U.S. counties (or about 8%). This anecdotal evidence stresses that it is not clear if public

managers see their agencies, like municipal governments, as being in competition with one another. This might explain the limited participation of public agencies in voluntary benchmarking clubs. Second, building from the first difficulty in following private-style benchmarking in the public sector, research concluded that voluntary adherence to benchmarking clubs is not enough to actually produce improvement in the performance of public health organizations (Barretta 2008, 364) and municipalities (Williams, 2005:219-220).

Occasionally, in jurisdictions where one level of government has constitutional power over another level of government, mandatory benchmarking systems have been implemented. This expanded the pool of participants. The availability of comparative data in mandatory systems, which is arguably the point in any benchmarking system, comes then from all municipalities in a jurisdiction: not from an unrepresentative small sample of like-minded municipalities. Additionally to the availability of comparative groups coming from the universality of data collection, is the standardization of measures (Higgins, 2005:448). Moreover, in mandatory systems, the number of participants is consistent. In voluntary systems, municipalities can opt out at any time. For example, Chapel Hill, Garner, Rocky Mount and Shelby decided to revoke their membership in the North Carolina's system at some point (Williams, 2005:99). Needless to say, mandatory systems are unheard of in the private sector, as no entity has the authority to coerce all firms in a sector to share data with each other.

Types of benchmarking

Many variants of benchmarking are applied in the public sector. Many typologies of benchmarking also exist in the literature. There are numerous typologies of benchmarking because different authors focus on different attributes of a specific branch of performance measurement.

Meszaros and Owen (1997:13) find four types of benchmarking: (1) internal, (2) competitive, (3) noncompetitive, and (4) world-class. Internal benchmarking involves a comparison between analogous operations within an organization. This would be the most frequently used form of benchmarking in the United States (Foltin, 1999:44). Competitive benchmarking makes comparison to what is seen as the best of the 'direct competitor(s)'. Noncompetitive benchmarking also includes comparison to other organizations that are similar but not considered to be direct competitors. World-class benchmarking is a very ambitious type of benchmarking not often found in the public sector. As its name suggests, it involves identifying and? comparing the operations of an organization to a world-class leader in a given service.

Ammons (1997:12; 2000:107) discerns three principal forms of benchmarking in the public sector: (1) corporate-style benchmarking, (2) targets as benchmarks, and (3) comparison of performance statistics as benchmarks. The first type of benchmarking either focuses on a single key process or on the details of a selected process; it tries to include a cautious analysis of elements deemed responsible for a program's success (Ammons, 2000:108). It is not often used in the public sector, as public agencies do not

pursue a single bottom-line, but rather many and often contradictory goals. The second type of benchmarking is more internal in nature. The targets established with this type of benchmarking would come from an inspirational idealized level of performance. In the United States, it would be more common in independent performance initiatives at the state and local levels. Usually, targets would be "(…) set arbitrarily rather than pegged to a level of achievement demonstrated elsewhere. In that respect, targets as benchmarks are rarely tied to a 'best practice'" (Ammons, 2000:109). The third type is more external in nature. Comparative benchmarking would be more attractive to governments that want an assessment of their performance on several fronts; it would be more extensive than corporate-style benchmarking (Ammons, 1997:13-14, 2000:109). Targets from that kind of benchmarking would lack the arbitrary character of targets obtained by other means and would be more attainable and defendable (Ammons, 1997:13-14). This type of benchmarking would be suitable to establish performance diagnostics (Ammons, 2000:109), to determine the effects of operational changes, to offer context that is media-worthy, and to comfort citizens that "someone is minding the store" (Ammons, 1997:14).

Hatry (2006) identified eight major types of benchmarking in the public sector. His types of benchmarking overlap with some of the categories presented above. Their labels are self-explanatory. The major types of benchmarking would involve different bases of comparative information from:

1. Performance in the previous period,
2. Performance of similar organizational units of geographical areas,
3. Outcomes for different workloads or customer groups,
4. Different service delivery practices,

5. A recognized general standard,
6. Performance of other jurisdictions,
7. Performance of the private sector, and
8. Targets established at the beginning of the performance period (Hatry, 2006:139).

The similarities between these typologies rest on the internal and external bases of comparisons used in benchmarking by public organizations. Although the benefits of internal benchmarking are recognized (Foltin, 1999:44; Hall, 2007:286), the normative general agreement in the literature is that external benchmarking realizes the full potential of performance measurement (Keehley and MacBride, 1997:77; HM Treasury, 2003:33; Bird *et al.*, 2005:19). Furthermore, external benchmarking, that is comparisons with other organizations, is more powerful and useful than internal benchmarking (Ammons, 1997:14), that is comparisons limited within an organization, usually historical performance trends.

One of the limitations of internal benchmarking, which is not shared by its external variant, would be the difficulty to assess the reasonability of targets and achieved performance (HM Treasury, 2003:33; Williams, 2005: 68-69; Behn, 2008:41). At the municipal level, there are recorded cases where isolation from external information resulted in weak and ineffective agencies, as later determined by the British Audit Commission, offering themselves as models of 'best practices' (Jas and Skelcher, 2005:204-205). To that extent, governmental agencies like HM Treasury recognize that an over-reliance on historical [internal] performance data results in agencies setting targets that "lack the necessary ambition to drive forward significant change" (HM Treasury, 2003:33). The same agency offers evidence of this behavior in local

governments from a governmental review: "only 30 per cent of the targets set for 1999/2000 aimed at an improvement over the level of performance achieved in 1998/99" (HM Treasury, 2003:33). Preceding Hinton, Francis and Holloway (2000:54), Keehly and MacBride (1997:77) observed that benchmarking has to skip over internal comparisons, otherwise "breakthrough improvement is impossible."

Benefits of benchmarking

Benchmarking is reputed to bring about certain benefits for organizations that include it in their operations. Similarly to the alleged benefits of performance measurement, benefits of benchmarking are identified by following different kinds of reasoning and ontological frameworks. On top of the inherent difficulties to pin down benefits (but also shortcomings) to the practice of performance measurement explained earlier, authors almost never mention from which type of benchmarking stem the benefits they observed or deemed present. Keeping this fact in mind is important while reviewing the benefits of benchmarking.

One of the alleged benefits of benchmarking is that it fosters transparency. By offering contextualization in the comparative data to stakeholders outside of an organization, benchmarking would make it possible for elected officials and citizens to pass judgment on the performance of their government (Siverbo and Johansson, 2006:275; Braadbaart, 2007:690). The empirical findings from Braadbaart's (2007) study of private providers of public services are informative about the effects of transparency. The author scrutinized the performance of 11 Dutch water supply utility providers that participated in a

benchmarking effort from 1989 to 2000. Changes in reporting in the benchmarking initiative happened in 1992 and 1997, when firms had to start to report their performance respectively to governing boards and the public at large. The results from this study are that prices charged to water users went down and converged, allegedly because of the benchmarking effort, but only after 1997 (Braadbaart, 2007:690). This is one of the few instances where it is hinted that improved performance can be achieved by increased transparency.

Most of the benefits related to benchmarking were identified from studies based on interviews of managers using this management tool. Holloway, Francis and Hinton (1999) came up with such a study. The authors used postal surveys to assess the perceptions of NHS General Managers and management accountants based in the U.K. What the authors found is that benchmarking is a way to uncover new ideas, and a catalyzing force for improvement from observed best practice (Holloway, Francis and Hinton, 1999:355).

A second perception-based study used focus groups and twenty-five semi-structured interviews of local managers in Northern Ireland (McAdams and O'Neil, 2002). From the interviews, the benefits identified by local managers under the Northern Ireland (pre-CPA) benchmarking system are that:

1. Benchmarking can help to improve the service by learning from other providers;
2. It can change the way of working,
3. It is a recognized process; therefore it has high credibility;
4. Before you can improve, you have to get to know where you are first;

5. For an organization that wishes to consider a quality-driven approach (e.g. the business excellence model), benchmarking is an essential element; and
6. It may help to meet business plan targets (McAdams and O'Neil, 2002:452).

The results from McAdams and O'Neil's focus groups in Britain are confirmed from the interviews of Bowerman and colleagues (2002:443) who found the same perceived benefits from practitioners.

A third study that identified benefits of benchmarking from interviews of managers was done in the North Carolina Benchmarking Project. Managers from that voluntary benchmarking project reported several perceived benefits from their experiences. Among other things, benchmarking would (1) enhance program accountability, (2) change the organizational culture, (3) foster organizational learning, (4) enhance the reliability of data, (5) facilitate communication within an organization, (6) provide assistance for budget decisions (Rivenbark and Ammons, 2007:40)

Benchmarking, just as other variations of performance measurement, is thought to drive performance improvement (Siverbo and Johansson, 2006:274-275). One of the most convincing studies demonstrating this benefit took place within KOSTRA, the Norwegian municipal mandatory benchmarking system. Revelli and Tovmo (2007) looked at the impact of whether municipal politicians compare their own municipality's performances in the provision of public services to those in other jurisdictions on the efficiency of their government's operations. What they found is that *ceteris paribus*, municipalities where elected officials use benchmarking data were more efficient than

municipalities that do not. Moreover, the authors factored in the spatial patterns of efficiency. Evidences of yardstick competition were found. Yardstick competition is that "informational spillover from the fiscal policies enacted in the neighborhood affects the beliefs of an imperfectly informed electorate with respect to the competency and honesty of their own government" (Revelli and Tovmo, 2007:122). More specifically, that an increase of 1 percentage point in efficiency brings an efficiency increase of 0.7 percentage points in adjacent jurisdictions actively benchmarking with neighboring governments, but does not affect governments who do not benchmark (Revelli and Tovmo, 2007:131). The authors conclude that "spatial interaction turns out to be negligible for those governments that are not engaged in yardstick comparisons while spatial auto-correlation is strong and significant for those governments that compare their bureaus' performances to those in other jurisdictions" (Revelli and Tovmo, 2007:131).

Barriers to benchmarking

Being a subset of performance measurement, benchmarking shares some of its impediment factors with it. However, the literature contains some barriers that are seen as being specific to benchmarking, like incrementalism (Drew, 1997:449; Davis, 1998:267). Some of the barriers to benchmarking are found more often in voluntary networks or benchmarking clubs. Among them are suspicions from unwilling managers (Davis, 1998:268; Lawton, McKevitt and Millar, 2000:15), the obstruction of politicians (Lawton, McKevitt and Millar, 2000:15) and the narrowness in the selection of partners (Holloway, Francis and Hinton, 1999:355).

More barriers have been identified with the help of survey research. From survey data of NHS General Managers and management accountants based in the U.K., the barriers identified in descending order of frequency, were "(1) difficulties in finding partners who were considered suitable comparators, (2) resource constraints (especially time), (3) access to other organisations, (4) staff resistance and (5) confidentiality" (Holloway, Francis and Hinton,1999:355). Another survey research of 140 American private sector firms in voluntary benchmarking clubs showed the extent to which the following barriers were seen as impeding in these firms' benchmarking effort (Drew, 1997). In order of acceptance comes:

1. Time needed to conduct study,
2. Reluctance of partners to share data,
3. Lack of senior management support,
4. Finding a project champion,
5. Lack of knowledge/skills,
6. Finding organizations to study,
7. Finding suitable staff and team,
8. Cost of conducting the study,
9. Fear of disclosing information,
10. Developing teamwork,
11. Resistance to outside ideas,
12. Lack of equipment and technology,
13. Legal or regulatory considerations. (Drew, 1997:438)

Some of these barriers like #1, #2, #6, by design, would not be observed in a mandatory benchmarking system.

There remain barriers to benchmarking in structured mandatory systems. Here are some difficulties identified from mandatory systems. Difficulties of surmounting accounting differences between state jurisdictions have been offered to explain why benchmarking has not been embraced more widely at the state and local level (Foltin, 1999:44). Others pointed out that some managers claim that their organizations are unique, so that comparisons with other organizations would be meaningless (Ammons, 1999:105). Commenting on the reluctance of some local managers to go for external forms of benchmarking, Ammons and Rivenbark (2008) suggested that managers limiting themselves to internal benchmarking would do so for defensive reasons:

> While comparison with one's own performance at earlier periods of time is important, reluctance to embrace external comparison is odd for a participant in a project designed primarily for that purpose and may reveal an underlying distrust of performance measurement, anxiety about the numbers being produced and what they will suggest about relative standing, or a lack of confidence in the organization's ability to improve performance. (Ammons and Rivenbark, 2008:311)

Another reason is more psychological in nature and does not assume "turfgarding" (Drew, 1997:429) on the part of managers. Quite simply, individuals are more knowledgeable and comfortable about the characteristics of their organization than those of another. "Internal characteristics also require less cognitive effort to recall and comprehend. For these reasons, internal reference points are particularly likely to be used when complex or difficult characteristics are being investigated" (Yockey and Kruml, 2009:101). Even when barriers are overcome, benchmarking does not provide only benefits: there are also shortcomings inherent to this activity.

Shortcomings of benchmarking

At the risk of being repetitive, many of the shortcomings of benchmarking come from performance measurement and are not specific to benchmarking thus they will not be replicated here. The literature does include shortcomings particular to benchmarking. What distinguishes the shortcomings specific to benchmarking from others associated with performance measurement at large is the ontological nature of the arguments: the perceived shortcomings of benchmarking present in the literature are observational and critical rather than empirically based. Once more, depending on one's preexisting proclivities for this structured and rational-positivist management tool, the paucity of empirically demonstrated benchmarking shortcomings will likely further antagonize proponents and opponents of performance measurement.

One of the perceived shortcomings originates from the top-down nature of mandatory benchmarking systems. Depending on the degree of centralization of such systems, measures and goals determined by a higher level of government can be seen by public servants as inflexibility and a negation of their professional judgment and discretion (Rogerson, 1995:28). A resulting lack of ownership would especially be felt (Loveday, 2006:292) for services with a strong organizational culture like policing. Imposed measures could aggravate employees of organizations that are not comparable with others. The benchmarking of parks in England is a quintessential example of problematic systematic use of performance measures. An Area Grounds Officer interviewed by Higgins (2005) had this to say:

> The performance side is important, but it needs to be done in context, and people have become besotted with where they are in the league, and we have a silly one

> on parks called ‘cost per hectare’ to maintain. Now some authorities have vast areas that they stick a few sheep in and that is it. Whereas here there are Royal parks which are very high on horticultural maintenance standards, very expensive in terms of price per metre. One of our central parks is perhaps thirty times that of a verge elsewhere. (Higgins, 2005:458)

Indeed, comparing local parks with Royal British gardens can render a distorted image of reality.

Another shortcoming of benchmarking would manifest itself in simple systems where a built-in referent to the mean is included in rankings. Assuming a normal distribution where the mean and the median would coincide, rankings based on comparisons with the mean invariably signify that half of the organizations are labeled as being *below average*. It does not take into account what the performance adequacy of the median/average organization is. Below average organizations would be compelled, and in certain benchmarking systems, forced to make operational changes that would not be warranted if inspectors would independently evaluate these organizations (Triantafillou, 2007:844).

The problem described above can be compounded if independently evaluated performance measure or service by service evaluations are components of a benchmarking system. Halachmi (2002a) offered three warnings about the consequences of possible shortcomings of benchmarking activities. The first two are relevant for the present analysis. The third one, which is not specific to benchmarking, will be eluded here, as it concerns the well-known discrepancy between citizens’ perception and actual performance.

> 1. First, the productivity of delivering any one service is a function of the total content of the basket of services offered by the local authority. Since various local

> authorities use different baskets and those baskets change over time, comparisons for studying any one service may be problematic. (…)
>
> 2. Second, the fact that the boundaries of any local authority are not optimal for delivering all the services it offers its residents. Thus, even when the area or the population served by the local authority allows the optimizing of one service this may at the expense of optimal performance when it comes to delivering one or more of the other services. Comparisons of a single service delivery among authorities could be misleading if its overall performance is not highlighted. However, measuring this overall corporate performance means finding measures for a series of governance achievements over and above service level achievements. At this time our measurement tools in this area are very weak. (…) (Halachmi, 2002a:217)

We acknowledge the plausibility of the shortcomings presented above by the author. However, as we discussed earlier, our measurement tool box is actually well stocked. It is possible to take into account the realities and complexities of budget choices. It is possible to evaluate simultaneously the performance of different services financed from a fixed resource pool. As we have seen in the *types of performance measures* section of this literature review, stochastic or frontier measurement tools like *free disposable haul* and *data envelopment analysis* have been around for more than fifty years.

All in all, we have seen in the previous sections that benchmarking is a specific way to measure performance. To borrow images from statistics, benchmarking is more like *time-series* and *panel data*, where as many independent performance measurement initiatives are more *cross-sectional* in nature. Although benchmarking shares many of the characteristics with more general forms of performance measurement, it also bears

specificities. Now that the theoretical foundations have been set, the next section will delve into an important facet of an information-based management tool: its use.

Use

Use of performance measurement

Prior to studying *sensemaking* of performance information, the *use* of performance information has to be addressed. *Use* is a necessary condition for *sensemaking*. In the case that interests us, local managers in Quebec who are not using the municipal management indicators at all, are not making sense of the indicators' values. Before reviewing sensemaking, the interpretative sub-phase contained in Daft and Weick's (1984:286) framework, between data collection and the taking of actions, we will cover the information utilization phase involved in the overall performance system (Lu, 2008:12).

It has been debated that performance information is not used by managers (Moynihan, 2008b:45, 67). Despite the fact that the volume of performance information grew substantially since the end of the 1990s, "overall [at the federal level] the use of performance information in management decision making has not changed over the last 10 years" (Steinhardt, 2008:2). There are different kinds of performance information use. The actual use of performance information excludes "(…) simply reporting measures or somewhat vaguely considering measures when monitoring operations" and implies the"(…) evidence of an impact on decisions at some level of the organization" (Ammons and Rivenbark, 2008:305). There are authors who suggest that looking at the actual (demonstrable) use of performance information neglects other more symbolic functions,

like showing that government cares about performance. Empirically, it is very difficult to identify where symbolic use starts and where simple compliance ends. A charitable empirical definition of symbolic use of performance information would be agencies participating in a performance measurement initiative, without evidence that the information was actually used. To that extent, de Lancer-Julnes (2006) offered a description of performance information usage beyond actual use:

> To those who would judge the effectiveness of performance measures in terms of their direct use in decision-making, the conclusion would therefore be that performance measurement does not have an impact on program management. But this interpretation would miss the more subtle, but still valuable, impacts of performance measures. Specifically, respondents reported using performance measures (1) to justify budgets; (2) to fulfill mandates for transparency (accountability as oversight and compliance); and (3) to inform debate among administrators and elected officials. (de Lancer-Julnes, 2006:227)

It is worth keeping in mind that studies assessing the extent of the use of performance measurement seldom take into account these subtle symbolic uses of performance information. The general use of performance information has been studied before for local managers (Wang and Berman, 2001; Chung, 2005; Rogers, 2006; Lu, 2007) and local politicians (Askim, 2007, 2008).

The consensus in earlier literature on the use of performance information was that it was seldom used. Summarizing the literature on the use of performance information in the public and private sectors, Feldman and March (1981) suggested the following reasons to explain the low level of usage:

1. Much of the information that is gathered and communicated by individuals and organizations has little decision relevance;
2. Much of the information that is used to justify a decision is collected and interpreted after the decision has been made, or substantially made;

3. Much of the information gathered in response to requests for information is not considered in the making of decisions for which it was re-quested;
4. Regardless of the information available at the time a decision is first considered, more information is requested;
5. Complaints that an organization does not have enough information to make a decision occur while available information is ignored;
6. The relevance of the information provided in the decision-making process to the decision being made is less conspicuous than is the insistence on information. In short, most organizations and individuals often collect more information than they use or can reasonably expect to use in the making of decisions. At the same time, they appear to be constantly needing or requesting more information, or complaining about inadequacies in information (Feldman and March, 1981:174).

In the past, when use of performance measurement was referred to, it was usually referred to in a general way, not unlike a census. In these studies, scholars tried to establish the proportion of local/state agencies that used certain kinds of measures. The typical findings are that output measures were more prevalent than outcomes measures (Usher and Cornia, 1981:233; McGowan and Poister, 1985:534; Jreisat, 1987:8; Palmer, 1993:32; Berman and Wang, 2000:413; de Lancer-Julnes and Holzer, 2001:699; Wang and Berman, 2001: 414). Other findings at the municipal level are that the use of effectiveness and efficiency was less prevalent in certain regions, like the southern cities (Lindblad, 2006:663). Contrary to expectations, the size of local government would not have an impact on the extent of performance measurement use.

The use of performance measurement, in these first generation performance measurement studies, was operationalized by scholars as the percentage of municipalities reporting

consumption of performance information in general terms like decision making (Fountain, 1997) or the use "in selected departments or program areas" (Streib and Poister, 1999:111) for processes like agency requests (Wang, 2000:107).

Bridging the first (input/output/outcomes mix) and second generation (task relying on performance information) use studies was Berman and Wang's study (2000). In their study of U.S. counties with population over 50,000, Berman and Wang (2000) tried to establish if organizations saying they use performance information would have the capacity to include this information in their operations. The authors found that among those that use performance measurement, about one-third had what the authors regarded as an 'adequate' level of capacity (Berman and Wang, 2000:417). By capacity, Berman and Wang (2000:417) meant that the counties met standard prescriptions from the literature, being able to: "(1) relate outputs to operations; (2) collect timely data; have (3) staff capable of analyzing performance data; (4) adequate information technology; and support from (5) department heads and (6) elected officials."

In second generation performance measurement studies, the use of performance measurement information concentrated less on the proportion of input, output or outcomes measures, and more on activities and functionalities where performance information was used (eg. control programs; allocate funds, motivate agency personnel; communicate the agency's programs to stakeholders, etc.). Using Berman and Wang's (2000) data, Wang and Berman (2001) assessed the link between what they called the 'deployment' and 'purposes' of output and outcome measures in county government. In

order to achieve that, the authors asked county managers if their "jurisdiction uses performance measurement to:"

1. Communicate among managers,
2. Communicate between managers and commission,
3. Communicate between public officials and residents,
4. Assess service delivery alternatives,
5. Force discussion about service problems and solutions,
6. Evaluate program results and achievements,
7. Monitor the efficiency/effectiveness of services,
8. Determine funding priorities across programs,
9. Determine funding levels for individual programs. (Wang and Berman, 2001:415)

The proclaimed use of performance information was quite high. It spanned from 53 percent for item #8, to 82 percent for item #2. Once more, we remind the reader that item #2 of the survey questions did not measure "actual use" of performance measurement (Ammons and Rivenbark, 2008:305).

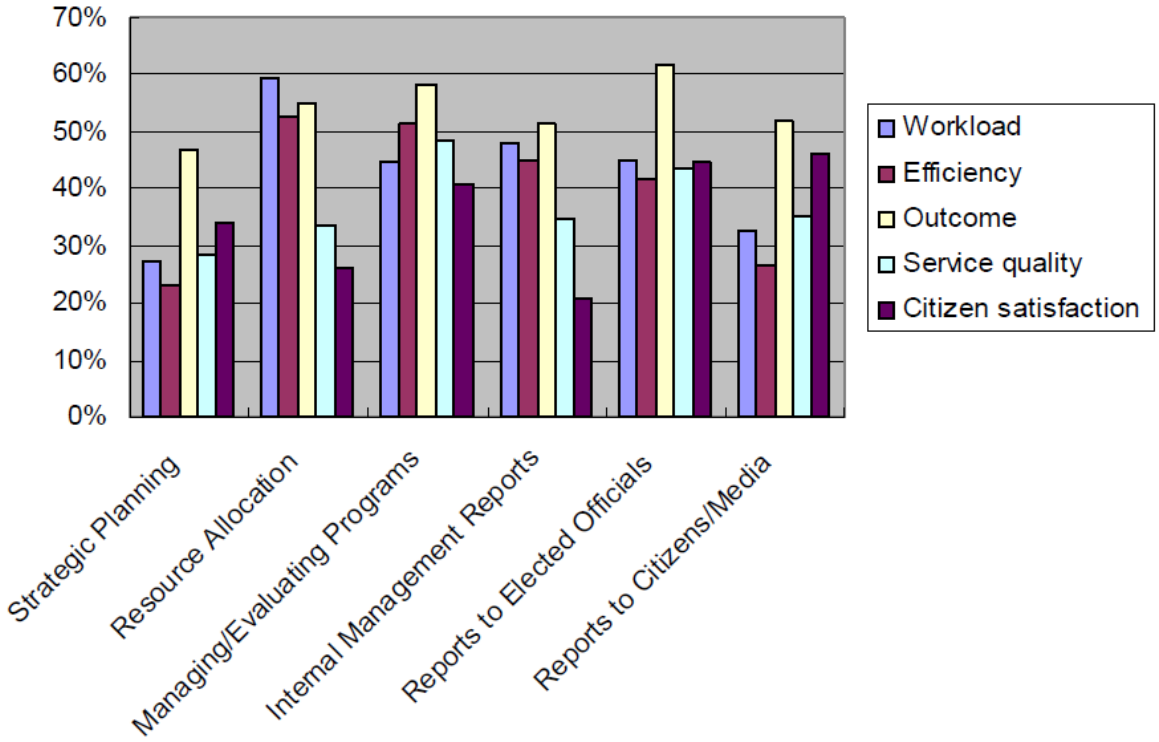
Willoughby (2004) used surveys from 212 budget officers from 36 states and 152 agency staff personnel from 48 states assessed the extent of performance information uses for various functions. The results were that budget officers and agency staff personnel agreed that performance information was *not* actually used for the following management functions:

1. To establish contracts for services;
2. To manage operations, such as scheduling activities;
3. To make personnel decisions regarding staffing levels and evaluations;
4. To evaluate underlying reasons for results;
5. To benchmark or compare program results with other entities;

6. To hold local jurisdictions accountable for state-funded or state-regulated programs;
7. For specific performance improvement initiatives such as technical assistance or Operations' improvements; and
8. To determine which programs, local jurisdictions, or contractors to target for audits, special studies, technical assistance, or other initiatives (Willoughby, 2004:31).

Slightly different results were obtained using ICMA data from managers in U.S. municipalities with populations between 25,000 and 250,000 (Chung, 2005). The results from 173 managers who reported having performance measurement initiative, were that performance information was reported to be used for approximately between 20 percent and 45 percent of municipalities, depending on the function (Chung, 2005:116). For a more detailed portrait of the reported use of performance information for strategic planning, resource allocation, managing/evaluating programs, internal management reports, reports to elected officials, reports to citizens/media, see table 3. One should keep in mind that the survey in question does unequivocally measure the actual use of performance measurement.

Figure 3. Types of Performance Measures Used for Management Activities



Source: Chung (2005:416)

At this point, it should be clear that establishing from the literature the proportion of local government managers disclaiming that they are using performance information is difficult to do in a precise manner. The samples and the questionnaires alone vary from study to study. For example, studies using ICMA data (Chung, 2005) have a different question formulation on performance measurement use than studies using GASB data (Rogers, 2006). Some authors studied peripheral concepts to use, like implementation of performance measurement (de Lancer-Julnes and Holzer, 2001:700-701). Some studies measure the use of performance measurement too generally, like when the GFOA allegedly asked jurisdiction if performance measurement was being used in their governments in 'some way' (Kinney, 2008:47). To add to the difficulties, some authors suggest survey questions to measure the span of the use of performance information, but

no do not use them or do not report descriptive statistics of their instrument, for example Jansen (2008:176). Two of the few studies which present their survey results in details are Rogers's (2006) and Lu's (2007).

Rogers' (2006) study used 277 GASB-generated surveys to assess the use of performance measurement in diverse local initiatives in the United States. The unit of analysis here is different than Chung's (2005) and Lu's (2007). "Use" in Rogers' (2006) study looks at the proportion of departments within a municipality that reportedly use performance information, whereas Chung (2005) looks at municipalities as a whole and Lu (2007) keeps track of the proportion of decisions for which performance information is used. Rogers' (2006) detailed descriptive statistics of the use of performance measurement are presented in table 3.

Table 3. Reporting Utilization of PM Output or Outcome Performance Measures

Extent to which PM used in:	Variable	No departments are using	A few select departments are using	Less than 50 % of departments are using	Greater than 50% of departments are using	All departments are using	Not Sure (System Missing)
Managing operations or daily decisions	Reported PM Utilization for Management	15.2%	26.4%	18.8%	16.2%	9.7%	13.7%
Personnel decisions including staffing levels and evaluation	Reported PM Utilization for Management	20.9%	26.4%	15.2%	12.3%	11.6%	13.7%
To hold local jurisdictions accountable for state requirements	Reported PM Utilization for Management	28.9%	17.7%	12.3%	9%	9%	23.1%
Establishing contracts for services	Reported PM Utilization for Management	18.4%	27.1%	11.6%	17.3%	5.8%	19.9%
Performance improvement initiatives	Reported PM Utilization for Management	19.9%	31.4%	14.4%	12.6%	8.3%	13.4%
Evaluations to determine underlying reasons for results	Reported PM Utilization for Management	18.1%	28.2%	13.4%	13.7%	9.7%	17%
To determine which areas to target for audits, special studies, technical assistance, or other initiatives	Reported PM Utilization for Management	35%	19.5%	10.8%	7.9%	5.1%	21.7%
Budgeting Decisions, including Resource Allocation or discussion about Resource Changes	Reported PM Utilization for Budgeting	15.9%	22.4%	13.7%	20.2%	19.9%	7.9%
Outcome/Output PM appearance in agency budget requests	Reported PM Utilization for Budgeting	14.8%	14.1%	9.4%	19.5%	34.3%	7.9%
Outcome/Output PM in the 1999-2000 Budget Document	Reported PM Utilization for Budgeting	19.9%	9.4%	6.5%	17.7%	36.5%	10.1%

Table 3. Reporting Utilization of PM Output or Outcome Performance Measures (continued)

Outcome/Output PM in the Annual operating budgets	Reported PM Utilization for Budgeting	15.9%	9.7%	6.9%	18.8%	40.4%	8.3%
Outcome/Output PM in budgeting quarterly reports	Reported PM Utilization for Budgeting	37.9%	14.4%	4.3%	9.7%	13%	20.6%
Outcome/Output PM in annual budget reports	Reported PM utilization for Budgeting	23.5%	18.1%	7.2%	15.5%	22%	13.7%
Reporting to management and staff	Reported PM Utilization for Reporting	10.8%	24.2%	11.9%	20.6%	24.2%	8.3%
Reporting or accountability to elected officials	Reported PM Utilization for Reporting	13.0%	22.4%	9.4%	19.9%	27.4%	7.9%
Reporting or accountability to citizens, citizen groups, or media	Reported PM Utilization for Reporting	18.1%	26.7%	11.2%	14.1%	19.5%	10.5%

Source: Rogers (2006:98-99)

Once more, the results point out that the reported use of performance information in the departments is rarely higher than 50 percent.

Lu's (2007) study targets state fiscal/budget officers in Georgia. The general findings are that performance information is used more for managerial purposes than for budgetary ones and somewhat more in budget increases/decreases requests than in redistribution requests (Lu, 2007:9). The study also points out that performance information is used to varying degrees by the Governor, the House and Senate Budget Offices and the legislature (Lu, 2007:11). The results of more targeted uses of performance information are featured in table 4.

Table 4. Use of Performance Measures for Specific Purposes in State Agencies in Georgia

To what extent do you (or you perceive other participants) use performance measures for the following purposes? (1 = *Never*, 2 = *<25% of decisions*, 3 = *25–50% of decisions*, 4 = *50–70% of decisions*, and 5≥*75% of decisions*)	Score given by state agency
By specific uses	**3.40**[a]
Determine performance	3.79
Improve performance	3.70
Benchmark data	3.63
Communicate programs to stakeholders	3.58
Improve strategic planning	3.47
Budget submission	3.45
Motivate agency staff	3.26
Control agency/program	3.18
Allocate funds	3.07
Reduce overlapping services	2.82
By funding decisions	**3.12**[a]
Budget enhancement request	3.18
Budget reduction request	3.16
Budget redistribution request	3.03
By participants	**2.76**[a]
Budgetary use by OPB	3.17
Budgetary use by Governor	3.07
Budgetary use by House and Senate Budget offices	2.67
Budgetary use by Senate Appropriation Committee	2.46
Budgetary use by House Appropriation Committee	2.45

[a]Mean of scores by that category.

Source: Lu (2007:10)

From this detailed table, some similarities can be observed with the Chung (2005) study. The most striking is that the proportion of performance information use lies between a range of 25 to 50 percent.

As for the use of performance information by local politicians, Askim (2007, 2008) offered precise studies. Using surveys of more than 700 municipal politicians in Norway, the author (2007:461) stated that 43 percent of surveyed politicians claim to regularly

consult performance information in municipal annual reports and balanced scorecard reports for decision making. A later study by the same author found that experienced or highly educated councilors tend to use performance information less (Askim, 2009:37, 40).

Hypotheses on the use of performance measurement

The present research tries to answer the following question:

R1. Which factors account for the uses of performance measurement by municipal managers?

From the literature, certain interrogations remain in respect to what influences the use of performance measurement by municipal managers. The following predictions will be tested:

H1. It is expected that managers who express their unwillingness to use performance indicators, will indeed use performance measurement less than managers that do not perceive this barrier.

H2. It is expected that managers who express their inability to use performance indicators, will indeed use performance measurement less than managers that do not perceive this barrier.

H3. It is expected that managers who express being prevented from using performance indicators, will indeed use performance measurement less than managers that do not perceive this barrier.

H4. It is expected that managers who describe their municipality as having clear priorities, will use performance measurement more than managers that do not perceive their municipality this way.

H5. It is expected that managers who describe their municipality as having internal working partnership, will use performance measurement more than managers that do not perceive their municipality this way.

H6. It is expected that managers who describe their municipality as having links between priorities and community needs, will use performance

measurement more than managers that do not perceive their municipality this way.

H7. It is expected that managers who describe their municipality as being under strong political leadership in regards to performance measurement, will use performance measurement more than managers that do not perceive their municipality this way.

H8. It is expected that managers who describe their municipality as being under strong managerial leadership in regards to performance measurement, will use performance measurement more than managers that do not perceive their municipality this way.

In the next section, we will cover the use of performance information that is particular to benchmarking.

Use of benchmarking

As we have seen, there are few empirical studies on the use of performance measurement information by managers and politicians. It should not come as a surprise that there are even fewer studies on the use of benchmarking information. One study previously mentioned in this research, Lu's (2007) analysis on the performance management behavior of state budget and financial managers in Georgia, also offered a glimpse on the use of benchmarking. What the author found is that budgeting and financial managers claim to use benchmarking information in less than 50 percent of decisions (Lu, 2007:9). One of the rare studies that do present their results on managers' use of benchmarking information is Johansson and Siverbo's (2009) study of 207 out of 290 Swedish municipalities. They found with their survey that 40 percent of Swedish municipal managers were reporting to use *relative performance evaluations*, that is, comparative benchmarking information, 'to a great extent' (Johansson and Siverbo, 2009:207).

A lot of what can be found in the literature on municipal benchmarking comes from the United Kingdom. It should not come as a surprise that a lot of what academics know about the use of benchmarking also comes from the different performance measurement and benchmarking studies from countries in the United Kingdom.

One of the earliest studies covering the performance management behavior in general and the use of benchmarking in particular, focused on local authorities in England, before the implementation of the Best Value system and the Comprehensive Performance Assessment system (Palmer, 1993). The author asked managers to identify if they used certain indicators as comparators, or benchmarks. Her results on the use of benchmarking do not specify what is meant by 'uses', what is the frequency of use, what is the percentage of decisions for which they are used, or if they impacted decisions. Still, it is informative to see that 63 percent of managers expressed that they used internal (historical) benchmarking and 56 percent indicated that they used external benchmarking (comparisons with other local authorities) (Palmer, 1993:33). In 1999, also in the pre-Best Value and pre-CPA era, postal surveys to General Managers and management accountants based in the UK showed that only a third of all respondents, when identifying the reasons for participating in benchmarking activities, "saw benchmarking as a source of new ideas, or route to improvement building on observed best practice" (Holloway, Francis and Hinton, 1999:355). Later, Boyne and colleagues (2002) performed content analysis of 'performance plans' in Wales, under the Best Value system. They paid specific attention to the presence of benchmarking information contained in these 'performance plans'. What the authors found is that:

> The percentage of plans including comparisons of performance is extremely low. This limited use of comparisons is surprising because bench-marking was one of the key elements of the review. Some plans contained comparative data gained through benchmarking, but not all pilots who were members of the same benchmarking club included the data. Some PPs utilized the Citizen's Charter indicators published by the Audit Commission. Only a few pilots produced extensive comparative data in the PP. In some cases comparative data are provided, but are difficult to interpret as there is little or no information on the comparator organizations. (Boyne *et al.*, 2002:703)

This low use of benchmarking data might have been one of the reasons that fostered the creation of complex mandatory municipal benchmarking systems in the UK: the national 'Comprehensive Performance Assessment' systems.

As we have seen in this section, the expression 'use of performance measurement' or 'use of benchmarking information' takes different meanings in different studies. By 'performance measurement use', some authors really mean performance measures collected. If the reader can tolerate some discrepancies between the few different perception surveys that studied the role of performance information by local managers and politicians and the inherent bias for respondents to overstate socially desirable management behaviors, an overall picture immerges. All in all, the literature points out at a relatively low use of performance information. We covered many of the reasons throughout this section; we will not repeat them here. However, one of the reasons for low use is the different usage behaviors of managers and elected officials. Put simply, politicians have different preferences than managers about their sources of information (Jansen, 2009:184). Students of public administration will not be surprised that once again, the politics/administration dichotomy endures.

Comparisons

Why comparisons?

For an information-based management tool like performance measurement, comparisons are the cornerstones on which the analyses of information are to be based. Comparative analysis, according to Askim (2008:125) constitutes the second of the three routinized activities of performance measurement. Furthermore, as we have seen earlier, comparisons are the prime difference between general performance measurement and benchmarking. More importantly for this research, during sensemaking activities performed by managers, comparative information offers referents with which the value of indicators can be interpreted. Comparisons are also necessary to identify best practices (Raaum, 2007:48). It is important to briefly cover the ramifications of available and used comparisons in performance measurement before addressing the intricacies of sensemaking.

Early on, while studying the methods of determining productivity of municipal government, Wilcox (1896:391) made it clear that comparisons are a necessary condition in assessing what was then called productivity (what is currently called performance). One of the reasons vouching for the importance of comparative information is that actors, organizations and individuals, are notoriously deficient in assessing themselves in relation to others, when information about others is missing. We covered earlier an example from England, where a grossly underperforming local authority council, as a result of isolation from other councils, had a distorted view about the performance of one of its services that it deemed a best-practice to be emulated (Jas and Skelcher, 2005:204-

205). A second example of the inaccuracies of performance self-assessment comes from the Department of Transportation in the State of Georgia. Executive managers, office head managers and district staff, were asked to guess in advance what would the satisfaction level of Georgians be, in regards to State highways. The results are that predictions of managers were often inaccurate, and that the predictions were pessimist, guessing that the public's satisfaction would be less positive than it actually was (Poister and Thomas, 2007:288). A third example of the limitations of assessment without comparative information comes from school principals in the State of New Jersey. In a survey, principals were asked to self-assess by answering the following question "How would rate yourself compared to all other public school principals in the state?" (Richards and Height, 1988:20). The full results are reported in table 5.

Table 5. Principals' Self-Rating (cumulative percentages)

	Top 1%	*Top 5%*	*Top 10%*	*Top 25%*	*Top 50%*
Elementary	6	32	72	93	100
Middle	15	52	70	91	100
High school	21	52	79	93	100
All principals	2	42	74	93	100

Source: Richards and Height (1988:20)

As we can see in table 5, about three quarters of the participants thought they belonged to the top 10 percent of principals in the State of New Jersey. This clear bias in self-assessment made the authors conclude "another argument for reporting indicators across

sites is that principals are not good judges of their own abilities" (Richards and Height, 1988:25).

Available comparative data act as referents, which are needed in sensemaking. Systematic comparative data, like external benchmarking data, enable judgments that are not solely based on opinions or values, but on factual measurements (Triantafillou, 2007:839). Evidences from the private sector are that the articulation of reference points helps to shape strategic behaviors on the part of managers and employees. Fiegenbaum and Hart (1996) state that:

> Ultimately, through the management of reference points and other supporting interventions, the actual mental maps and schemata of organizational members might be shifted in the desired direction. At this point, the new behaviors have become internalized and no longer require direct management attention-they have become part of the identity and culture of the organization. (Fiegenbaum and Hart, 1996:232)

The limited literature on comparisons in municipal benchmarking systems motivates the present research to devote attention in answering the following question:

> **R2. What are the comparison levels being used by municipal managers in interpreting performance measures?**

In regard to this research question, we make the following prediction:

> **H9. It is expected that, despite the availability of external comparative data, the most frequent comparison level used by managers will be historical comparisons of their own municipality.**

All in all, comparisons in performance management help managers make sense of performance information and offer guidance to subordinates on the satisfactory level of performance that is sought.

Standards

As we have seen earlier, there is a debate about what constitutes the optimal level of comparisons and benchmarking. Without duplicating previous descriptions here, suffice to say that it is sometimes mentioned that external benchmarking is superior to internal benchmarking (Keehley and MacBride, 1997:77; HM Treasury, 2003:33; Bird *et al.*, 2005:19). Similarly, keeping track and reporting the value of indicators for multiple years is deemed superior than only providing performance for the current year (Coe, 2003:58); the same is said for the presence of targets accompanying the values of indicators (Ho and Ni, 2005:75). One decisive aspect of benchmarking that has not been covered yet is the debate about the presence/absence of available comparative baseline and the presence/absence of standards.

By standards, we borrow heavily from Miller and Miller's (1991:504) definition of standards in the context of citizen satisfaction to offer a definition for performance measurement and benchmarking. Standards are norms that offer clear prescriptions about what constitutes acceptable thresholds of performance.

Numerous descriptions on standards exist in the literature: they are sometimes compared to goals (Askim, 2008:125) or targets (Ritcher, 2004:19), treated as a special kind of reference point (Yockey and Kruml, 2009:97) or referred to as minimal thresholds (Bevan and Hood, 2005:9). In the context of public roads, Ritcher (2004:19) explained that "a performance standard can serve not only as a "target to shoot at" but also as a benchmark against which success can be assessed." The author adds that to be effective

in management, standards should be "well above average, but within the bounds of what has been achieved with current best practices and technologies" (Ritcher, 2004:19). Hence, standards are relevant for sensemaking, as they provide managers with a normative point to assess performance (*see* Talley, 1986:205) and initiate a course of action, corrective or not.

In 1927, Ridley commented on the existence of standards in the public sector. One of the earliest specialists of performance measurement commented that:

> The results of government are measurable. But the development of measurement standards will be a slow process. All sciences have developed slowly. The administration of government will become a science only to the degree that the results of government activities are subjected to measurement. Physics has been evolved from natural philosophy, astronomy from astrology, and chemistry from alchemy only as measurement standards have been applied to them. It will be so with government. (Ridley, 1927:47)

More than eighty years later, Ridley would notice that performance standards are still lacking. The absence of measurement standards is not limited to public administration; as Kent (2009:56) puts it, "developing such standards is the purpose of calibration, which is necessary and routine in the natural sciences but is not common in the social sciences." Obviously, calibration in the social sciences requires that at the very least, prior to calibrate a measure, a consensus be reached in the definitions of the measure itself. Taking another step back, a concept, or in the case that interests us, a public service, has to be measured. As we have seen earlier, many services in many jurisdictions still escape measurement. In the public sector, a conventional wisdom subsists that there are some things that just cannot be measured. The consequences are that,

> By taking this posture, people fail to get agreement on what is to be delivered and how accomplishment will be measured, where they are headed, and how to tell

> when they have arrived. This will allow you to argue that what you are doing and delivering is worthy, because there are no criteria by which to calibrate success or even to determine if progress is on target. (Kaufman, 2006:9)

Although this explanation could hold true, commentators would reply that the political nature of governmental services assures that even if such standards could be reached at some time, they would be likely to be transient and change when new political masters comes in.

A crucial aspect of performance measurement and benchmarking is that there is no obvious baseline to use while comparing performance information. Especially if standards are developed internally, it is difficult to assess if they are substantially calibrated. Internally developed standards can be theoretically difficult to build (Talley, 1986:209); as a result the so called standards are only the traditional way in which the agency operates. There are very rare examples in the literature where conscious and concerted efforts were deployed to come up with standards that were not linked with comparative data, but that rather came from managers and public servants (Pritchard *et al.*, 2009:88), and citizens (Smyth, Watzin and Manning, 2007:306). These are exceptional cases.

A plausible explanation about the lack of widespread standards in activities was offered by Schweigert (2006). For standards to exist, “the state of knowledge that informed the standards is assumed to be adequate so that better performance on the established scale is assumed to represent greater effectiveness” (Schweigert, 2006:424).

Wadman and DeLadurantey (1984) offered an explanation in the context of policing: "No consensus exists regarding what constitutes low, medium or high levels of crime. To establish such standards would require the accurate measurement of public tolerance, and this varies from region to region and from community to community" (Wadman and DeLadurantey, 1984:228). Without agreed-upon baselines, different types of comparisons are used. In the absence of any 'gold standard' for baseline, evidence will come from some type of comparisons, including rankings (Bailey and Hewson, 2004:512) and the different types of benchmarking we reviewed in a previous section.

Standards are a subcategory of comparisons that are more normatively charged than other types of comparisons and benchmarks. However, one should not lose sight that even objective standards, when they do exist, are often derived from the collective experience of many organizations: the case was made for municipal finances (Sohl *et al.*, 2009:75). This might be why benchmarking information is more readily available for sensemaking activities than are standards.

Target setting

Targets

Targets could be defined as "desired or promised levels of performance based on performance indicators. They may specify a minimum level of performance, or define aspirations for improvement" (Public Administration Select Committee, 2003:6). Accordingly, targets can serve as one of the comparison levels from which sensemaking activities can be performed regarding the analysis of performance information. Besides,

at a more basic level, targets can also serve as a powerful vehicle for transparency and performance improvement. Recommendations from the professional literature, but also the academic literature, regularly address the performance improvement and the transparency functions of targets. The normative assumption is that the presence of targets in performance measurement should be pursued, as it is superior to unaccompanied indicators (Ammons, 1997:14).

Among the foremost recommendations of governmental agencies are the findings of the British Public Administration Select Committee's (2003) "On Target? Government by Measurement" report to the House of Commons and Her Majesty's Treasury's (2003) "Setting Key Targets for Executive Agencies: A Guide". According to HM Treasury, the purpose of targets is to send out a signal of what organizations are trying to achieve and to prioritize the outcomes/outputs that are pursued. As such, according to HM Treasury, targets should:

1. [They] provide ambition and a sense of direction, concentrating efforts and resources on delivering the things that are important, and communicating clear messages to delivery staff.
2. [They] provide a focus on delivering results. Good targets should drive agencies to perform effectively and to deliver the key outputs and outcomes that underpin the aims of the organisation.
3. [They] provide a basis for monitoring performance. By stating what you are trying to achieve, and by tracking how you are doing, you can make judgments about how well your organisation is performing. (HM Treasury, 2003:9)

HM Treasury suggests that the targets themselves should satisfy a number of specifications to fulfill the functions that are presented above. Like many other sets of

recommendations from the professional literature, the characteristics of targets fit neatly in a mnemotechnic acronym: S.M.A.R.T.

1. Specific - so it's clear what it is that you are aiming to achieve;
2. Measurable - there should be a clear and transparent measure of success;
3. Achievable - the target should be stretching, and reflect the Government's ambitions for improved standards of public services. However, it must be achievable. Preferably there should be some evidence that demonstrates what is possible (e.g. benchmarking with similar organisations).
4. Relevant - the target should reflect what the organisation is trying to achieve - not simply what is easily measurable.
5. Timed - it should be clear when the target should be delivered by. (HM Treasury, 2003:12)

After extensive consultations with public managers, the Public Administration Select Committee recommended that the five aspirations of public sector agencies in regard of their targets should be that targets have:

1. a clear statement of what the Government is trying to achieve;
2. a clear sense of direction and ambition;
3. a focus on delivering results;
4. a basis for what is and is not working; and
5. better accountability. (Public Administration Select Committee, 2003:3)

The Ministry of Finance of Finland, in its comprehensive *Handbook of Performance Management*, also put together guidelines of general requirements for good performance targets. According to this publication, targets should be:

1. connected with and derived from the basic tasks of a government agency or operational entity (not random or individual);
2. strategic, essential from the point of view of operations and effective;
3. clear, understandable to everybody;

4. concrete, operational (practical from the point of view of monitoring);
5. evaluable, measurable and time-dependent;
6. preferably comparable (over time on the one hand and between similar units on the other);
7. such that the government agency itself can influence them;
8. realistic but challenging;
9. acceptable (jointly outlined and agreed);
10. such that they cover as large a part of the operations as possible while on the other hand determining priorities. (Salminen and Viitala, 2006:29)

The same research body also put forward that targets, to be effective, require a starting point and a goal. Only then would it be possible to make a diagnostic:

"targets require a starting point as well as a goal. Without knowing where things stand at present, there is no way of determining whether a target offers an organisation an easy goal, a challenge, or a target as remote as the moon" (Public Administration Select Committee, 2003:9).

Target setting

With some exceptions (Bird *et al.,* 2005:7-8), the literature on target setting does not often take into account what we covered earlier in the *non-linearity* section; the same could be said for decision makers in some agencies (Loveday, 2006:286). The literature is also very limited on practical knowledge on how targets are supposed to be set somewhere between current service levels (Roche, 2008:336) and pie-in-the-sky unachievable levels. Setting targets below the previous year's actual level, especially without explanation, is considered poor target setting (Ellig, 2007:6). A clear distinction exists between targets based on aspiration and targets extrapolated from evidence

(Marsden and Bonsall, 2006:194), usually practices collected through external benchmarking. To that effect, motivational targets that are rationally-based could be demoralizing because they are unachievable (Bird *et al.,* 2005:9).

There are few practical suggestions in the literature on how to set targets. For inspirational targets, one of the few recommendations to that extent was offered by Ammons (2000), when discussing State-level inspirational target setting. The author suggested that participation and deliberation via "(…) efforts to gain broad acceptance of that vision, and solicitation of cooperation by all parties who can help reduce the gap between current conditions and those desired" (Ammons, 2000:109) should be sought. Straight (2000:498-499) also suggests to involve managers and staff by techniques like brainstorming and multivoting, to maximize the chance that the targeted goals will be achieved. Myhre (2008) probably offered the most detailed and practical prescriptions of how to implicate employees in target setting. The author's recommendations are that:

> To ensure constructive participation, ascertain whether the measurement and evaluation scheme is perceived as valid, fair, and forgiving when necessary. Employees must agree that the measures fairly represent the results of their work in order for them to take ownership. If the performance measurement effort stresses performance against targets too soon, employees may challenge the data or how the targets were set, or worse they may engage in game-playing to set targets that will always be met. (Myhre, 2008:31)

For evidence-based targets, it has been advised to use the stochastic method that can take advantage of best practices, inherent constraints and non-linearity. Clarkson and Challis (2006:474) opined that "only the DEA [data envelopment analysis] method attempts to measure efficient performance with reference to the best performing (under current

constraints), and it is the only method able to offer targets for authorities to improve their practice."

The paucity of the literature on target setting in municipalities motivates devoting attention to the following question:

R3. How are targets set by municipal managers?

From this research question, the following prediction is made:

H10. It is expected that, despite the availability of external comparative data, the most frequent comparison level used by managers to set targets will be historical comparisons of their own municipality.

Targets for transparency

Targets can also contribute to increase transparency in performance reporting. Reporting targets alongside indicator values, because targets are easily understood, enables stakeholders outside a public agency to ascertain, at the very least, if there is a gap between a preexisting target and its indicator (Marsden and Bonsall, 2006:200). The presence of a target represents an improvement to the reporting of an unaccompanied indicator. Targets have strong advocates supporting their presence in performance management (Salminen and Viitala, 2006:24). Their argument is that preset targets are the *sine qua non* yardstick by which performance can be assessed. As Wall and Martin (2003:497) put it, "without this [targets], the whole exercise is meaningless and the [outside] user is just left with a number of inconsequential statistics". Nevertheless, the lack of clearly explained links between inputs, processes, outputs, outcomes, and overall agency strategy, even in the presence of targets, has been described as "(…) reporting of performance data rather than of performance assessments—an inadequate conversion of

data to information" (Tilbury, 2006:57). This is corroborated by other authors (Straight, 2000:499; Wall and Martin, 2003:497).

Targets for performance improvement

As a vehicle for performance improvement, targets would motivate organizations as a whole to achieve the level of performance identified by the target. It is taken for granted that targets for performance improvement should "(…) contain an element of stretch and ambition" (HM Treasury, 2003:33). It is firmly established that challenging goals and targets improve performance in a plethora of settings (for a convincing meta-analysis of the literature, see Locke and Latham, 2002:714). Practitioners and academics are usually quick to point out that targets should not be set at a level so high that it is unrealistic, unachievable (Matthews and Endress, 2008:131), and possibly demoralizing (HM Treasury, 2003:33). Surveyed managers of forty-three government departments in Australia and Hong Kong expressed their dissatisfaction with unrealistically set performance targets and their frustration with targets unmet as a result of factors outside the control of their organization (Taylor, 2006:342). There are empirical evidences supporting the claims that the presence of targets can increase performance. In the field of education, panel data from all 147 English local education authorities between 1998 and 2003 were analyzed to determine the influence of targets on student test scores (Boyne and Chen, 2007). The results were that "authorities with a target performed better than their peers in the LPSA period and better than themselves in the pre-LPSA period" (Boyne and Chen, 2007:472). A second study of 374 municipalities in England found a significant difference in the use of targets between "excellent" and "good" local

authorities, and “fair”, “weak” and “poor” ones (Boyne and Enticott, 2004:14). The latter would be less likely to agree that they have a “well-developed framework of performance targets to drive what we [they] do” (Boyne and Enticott, 2004:16).

Target for sensemaking

Targets can provide built-in threshold levels from which sensemaking can be readily performed. It is often assumed that targets, especially if they were grounded in best practices or some standards, become a clear demarcation below which performance is sanctioned and above which performance is rewarded (Courty, Heinrich and Marschke, 2005:324). Such routine institutionalized sensemaking would then be linked with performance improvement and transparency. This is not to say that targets can insufflate an absolutist approach to performance measurement where “(…) a target is either achieved or not achieved, real discussion about performance improvement may be suppressed” (Tilbury, 2006:52). However, the previous position is a minority take on targets. A more widespread view is that predetermined targets would be the yardstick to determine if a given performance level can be labeled and treated as a success. As such, meeting targets is often referred to as success (Kassel, 2008:242).

Empirically, Matthews and Endress (2008) in the field of environmental regulation and protection, elaborated on the sensemaking role of targets. On the relation between targets and sensemaking, the authors stated that “judgment of success or failure in compensatory mitigation wetlands is ideally based on goals established a priori” (Matthews and Endress, 2008:131). In a regulatory setting, sensemaking and target setting become very

important, because corrective actions and sanctions will be put forward according to performance verdicts. From the wetland protection example, improperly set targets transpose into sites that fail to achieve compliance (Matthews and Endress, 2008:131). The number of targets also becomes crucial, as for example, wetland sites with more targets, have been, ceteris paribus, designated as less successful than sites with fewer targets, mainly because it becomes more difficult to meet many targets than fewer targets (Matthews and Endress, 2008:134).

Despite their many transparency performance improvement and sensemaking functions, targets are not without critiques. First, many authors are fast to point out that a rigid top-down target regime is the first and foremost influence responsible for malign gaming. As we mentioned earlier, gaming has real costs and is very difficult to identify (Bevan and Hood, 2006:420-421). Second, targets can trigger benign gaming in the form of ratchet (*see* Robson, 2005:141) and threshold effects. The ratchet effect refers to the "tendency for central controllers to base next year's targets on last year's performance, meaning that managers who expect still to be in place in the next target period have a perverse incentive not to exceed targets even if they could easily do so (...)" (Bevan and Hood, 2005:9). The threshold effect refers to the "perverse incentive for those doing better than the target to allow their performance to deteriorate to the standard and more generally to crowd performance towards the target" (Bevan and Hood, 2005:9). It is hypothesized that in the presence of targets, efforts to limit the ratchet effect would increase the threshold effect, and vice versa (Bevan and Hood, 2005:9).Third, at a more general level, some critiques will borrow from monetary macroeconomics the so-called Goohart's Law

(Chrystal and Mizen, 2003:223), that is "when a measure becomes a target it ceases to be a valid measure" (Gutiérrez-Romero, Haubrich and McLean, 2008:784).

All in all, the presence of targets adds another layer of rational-positivism to performance measurement. As such, it exacerbates the potential benefits and drawbacks of this management tool, as it goes forward on the assumption that public sector service performance is well understood to the point where it can not only be monitored, but also steered (*see* Hoogenboezem and Hoogenboezem (2005:576) for examples in the field of policing). In the next section, the literature on behavioral patterns in meeting targets in the public sector will be reviewed.

Having covered basics of performance measurement and benchmarking, use, comparisons and target setting, we are now ready to cover the last part of the literature review: sensemaking.

Sensemaking

Data do not speak for themselves: analysts have to do it on their behalf. Making sense of the data is needed before actions are taken. As Behn (2008:212) put it "When the data speak, they do so only through some framework, some theory, some causal model, some logical construct, some perception of the world and how it works. After all, any set of data are just abstract numbers until they are connected to some reality (...)". A proof of that is that different individuals reach different, sometime polar conclusions from the same set of data: citizens, special interests groups and public managers in the society at

large; finance and human resources departments within organizations; and electors and politicians of different parties in the political arena (Moynihan, 2008a:25).

Weick (1995), in his seminal book *Sensemaking in Organizations*, offered this definition of sensemaking: "Although the word *sensemaking* may have an informal, poetic flavor, that should not mask the fact that it is literally just what it says it is" (Weick, 1995:16). In this research, adapting from Moynihan's description of interpretation (2008b:104), sensemaking is defined as the interpretation of whether performance is satisfactory in light of previous performance, performance achieved in other jurisdiction, professional standards, or some target, implicit or explicit. Moynihan (2008b:6) stated that performance management systems provide public officials with the structure to "(…) engage in coding - interpreting and refining information from the external environment and internal stakeholders into a series of information categories such as strategic goals, objectives, performance measures, and targets." Behn (2008:21) corroborates this idea that performance assessment is possible only through comparisons. Only after this interpretative phase can the information be useful to be presented to decision makers.

It has been argued that "given the basic nature of sensemaking—that it is improvisational and specific to local circumstances (..)" it is impossible to prescribe a "(…) comprehensive toolkit for managers to use in their sensemaking activities" (Jeong and Brower, 2008: 225). Although there are inherent difficulties to providing universal administrative prescriptions on such a complex topic, understanding the practice of

sensemaking in public sector performance measurement activities remains an important goal.

Sensemaking in the public sector is inherently complicated by the political nature of public administration. When studying the Norwegian benchmarking network, Askim, Johnsen and Christophersen (2008:306) observed that political conflicts in municipalities make it less likely that actors will interpret past experiences similarly. The absence of shared interpretation of past results limits the impact of benchmarking on future preferences and organizational behavior (Askim, Johnsen and Christophersen, 2008:306). Even within an agency, different actors have strategic interests to give meaning to performance information. To this effect, when offering five strategies for the use of performance measurement, De Bruijin (2002: 587) specified that a consensus between managers and professionals would be preferable to a monopoly on meaning giving. De Bruijin (2002) pointed out that even if managers are more removed than professionals from the reality of what is being measured, the meaning offered by professionals cannot be assumed to be always superior to the meaning given by managers. This is because "(…) professionals may have an interest in veiling their poor performance by assigning a particular meaning to it" (De Bruijin, 2002: 587).

Managers, when assessing performance information, can use different levels of sophistication in the sensemaking phase. Dutton's (1993) dichotomous categorization of managers' strategic diagnostics fits well with the definition of sensemaking that was presented earlier. Dutton (1993: 341) first identifies *automatic diagnosis*, which "(…)

involves the activation of ready-made issue categories in the minds of decision-makers that have been built from encounters with issues in the past". This relates to the referral of previous experience that was enumerated in our definition of sensemaking. Dutton (1993:342) also identifies *active strategic issue diagnosis*, where the selection of referrals is, in her own words, "(…) intentional and conscious, involving a much greater degree of information search and analysis". This search "(…) beyond the information that is readily apparent" (Dutton, 1993: 342) encompasses the other referrals of the sensemaking definition: *performance achieved in other jurisdiction*, *professional standards*, and *targets*.

In her landmark article on the scope and usage of performance information by local management in pre-Audit Commission Britain, Palmer (1993) pointed out that:

> (…) [I]n practice, the complexity of the data, the lack of guidance as to how to use or interpret the information and the problems associated with the use of performance indicators as comparators mean that few of these external decision-makers have the ability to draw meaningful inferences from the indicators (Palmer, 1993:34).

The importance of sensemaking in making a set of indicators useful for decision-makers has been recognized outside of municipal management. In their study of shared public sector oversight of water quality of Lake Champlain, Smyth, Watzin and Manning (2007:302) observed that the usefulness of a set of indicators for decision-makers depends on "(…) defining acceptable levels for each indicator so that the monitoring data that is collected can be interpreted". They added that without referrals about acceptable levels, performance information simply cannot be used for management decisions purposes. They emphasized that only by "(…) systematically separating acceptable

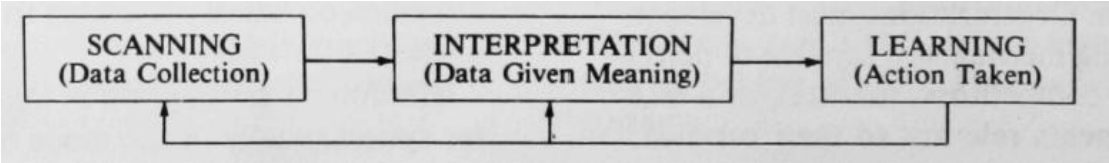
measured values of indicators from unacceptable values can managers identify those ecosystem components that require management attention" (Smyth, Watzin and Manning, 2007:302). Hatry (2008:227) identified that "(…) users can interpret the extent to which the measured levels of performance represent good or poor outcomes" as one of the "(…) five key "technical" elements that seem necessary for successful use of performance measurement information" (Hatry, 2008:226).

As we will present further in this research, the implicit or explicit assessment by managers of whether achieved performance levels represent a success or a failure, or are satisfying or unsatisfying, is imperative for public administrative scholars in order to understand how public managers use performance information to make decisions.

Success and failure

Success and failure are two of the verdicts that can be achieved by a manager trying to assess the performance information that is provided to him/her. As we have seen earlier, performance assessment, making a normative judgment about performance, is a critical and understudied phase in performance management. It is worth reproducing Daft and Weick's overall framework introduced at the beginning of this research, and reiterating Jeong and Brower's (2008:243) second stage of sensemaking, which is the framework delimiting the study of sensemaking in the present work.

Figure 4: Daft and Weick's Overall Framework where Sensemaking Lays



Source: Daft and Weick (1984:286)

Even though the public sector is not driven by competition and a bottom line, as long as there is some sort of agreement about what an agency is trying to accomplish, it is possible to distinguish a successful from an unsuccessful program (Hall, 2007:283). Moreover, "success and failure in public programmes is rarely so clearly defined (…), so governments (and analysts) have to understand and balance a range to say that an outcome is a success or failure" (Bovens, Hart and Peters, 2001:654). Methods to determine good and bad performance are to this day underdeveloped (Clarkson and Challis, 2006:462). Even in sophisticated and seemingly objective systems like the (now) defunct *Comprehensive Performance Assessment*, judgments about what a good performing local authority should be were added to the arithmetic average of the ratings point system determining 'excellent', 'good', 'fair', 'weak' or 'poor' local authority (Haubrich and McLean, 2006:95).

As if this was not enough, the time necessary to demonstrate success for public services is not in step with regular reporting (Davies, 2004:34). To add to the complexity, the definition of what constitutes success and failure, in the public sector, evolves with time (Hall, 2007:284). Changing definitions of success is not only influenced by technological advances and changing citizen expectations, but also by a changing mission. To convince ourselves, we can think about the example of a welfare program. Originally, a welfare

program is established to provide aid to the impoverished. A successful agency would be one that attains the greatest coverage of the targeted population. According to Hall (2007:284-285), this is precisely what happened: "the level of success began to breed a perception of failure over time, such that the welfare program was perceived as wasteful, with benefits going to undeserving persons". Eventually, when welfare was thought of in terms of self assistance through employment rather than government assistance, extended coverage was a sign of failure, not success (Hall, 2007: 285). Additionally, because there are many different dimensions to performance, there are different dimensions to success and failure. What to make of a service that meets some, but not all objectives? Again, sensemaking comes into play[**].

Success

An interpretation of a given performance level as satisfactory is important, since it is likely to validate and reinforce the strategy and approach that an organization is pursuing (Palmer and Short, 2001:212). Before passing review of what exactly is success in the public sector, we are reviewing what success is in the private sector, as the latter is more straightforward than in the former.

[**] There are examples, mainly from practitioners, where automatic diagnostic of performance was put together. The following example from Coryn, Schröter and Hanssen (2009) is informative:
Based on feedback from these three stakeholder groups, our conceptual definition of success for the case scenario presented here is threefold: (a) long-term retention of economically self-sufficient employment, (b) stability of housing, and (c) reduced need for government or other social support systems. Combined, these three objectives then became the criteria used to identify and classify success and failure cases, where a ''success'' case would have positive outcomes across all three objectives, an ''average'' case would have positive outcomes on one, or perhaps two, of these objectives, and a ''failure'' case would demonstrate poor outcomes across all three (Coryn, Schröter and Hanssen, 2009:83).

Success in the private sector

It is acknowledged that sensemaking is necessary in the private sector (Palmer and Short, 2001:211), despite the fact that definitions of success would be less debatable in the private sector than in the public sector. Private sector organizations, unlike public organizations, have a bottom line. "For them [private organizations], success is measured ultimately by profitability and the various elements that influence profit. Government performance has no single criterion of success" (Berman, 2008:3). This idea that success has clear cut measures in the private sector, like profit, performance-financial return, and market share; but not in the public sector, is found often in the literature (Allison, 1992: 285-286; Lonti and Gregory, 2007:480; Scorsone, 2008:61-62). Reviewing the literature of what constitutes success in the private sector, Palmer and Short (2001) uncovered that:

> The most common benchmark found in the literature, however, is industry average performance. In investigations of the effects of performance on strategic action, researchers routinely operationalize the referent construct as average return on assets (ROA) or equity (ROE). The assumption is that industry averages of these ratios are used by managers to determine whether their firm's performance is satisfactory (above the mean) or unsatisfactory (below the mean). (Palmer and Short, 2001:211)

This has at least two important ramifications. First, the most common benchmark for success would not be the presence of a profit, or more basically, that a firm survived another quarter (Ormerod, 2005:120). Obviously, the survival of a public organization is not an operational success and is never part of its mission (Eitel, 2004:248). However, survival remains a useful measure of success for private firms with nonoperational goals that are akin to public organizations (schools, colleges, hospitals) (Wilson, 1993:157). This criterion does not apply for public organizations, as their survival depends on the will of the political masters.

Success would be a function of the mean return on assets or equity for a comparable industry. Thus success would be the difference between one's return on assets or equity and the average performance of competing firms in the same industry. In the case of a normally distributed performing industry, where the mean, median and mode would be the same, virtually half the industry would be successful, while the other half would not. Next, it would mean that in the private sector, comparisons for sensemaking purposes are disconnected from comparisons for performance improvement.

Second, as we have seen, private organizations would use best practices comparisons, often across industries (which means with firms they do not compete with), when wanting to improve performance in some aspect of their operations. Meanwhile, firms would set back on comparisons with the industry's average in their sensemaking activities. For close to half the firms in an industry that are "successful", comparing themselves with an average would color sensemaking activities. This makes Palmer and Short (2001:215) say that "management at successful companies is biased to rely on industry averages when judging their performance because such indicators bolster self-worth and validate past strategies. Unfortunately, if unchecked, the practice can trigger the onset of organizational decline." Evidences from long term histories of 36 companies found this phenomenon to be true; former successful companies have a somewhat lower probability of being successful in the present period than unsuccessful, even mediocre counterparts (Miller, 1994:344). The author suggests the following reasons to explain this counterintuitive finding: "It seems that after success, organizations are indeed less able or

less inclined to adapt their structures and processes to changes in environment than after mediocre performance" (Miller, 1994:342).

Success in the public sector

Arguably, the most astounding characteristic of the notion of success in the public sector is that it is seldom defined. This is not to say it is not mentioned and referred to in the literature. It is. An example is provided by Willoughby (2004:23), while discussing success: "(…) practitioners and academics recognize that ''success'' (however defined) resulting from any performance measurement system often is conditional at best." In defense of academics and practitioners, success is difficult to ascertain in the public sector for many reasons. First, there is no bottom line for many agencies, just like there is no common bottom line for agencies across industries. A retail store chain and a restaurant chain *could* be compared on their return on investment; whereas there is little communality (if any) that could serve as a basis of comparison between fire stations and public schools. Second, success has meaning in a context, that is, in someone's context. Stakeholders of a publicly traded firm are owners; stakeholders of a public agency are not only service users, but the citizenry at large. Third, public sector organizations' performance is evaluated in terms of efficiency, effectiveness to objectives and a mission, and equity; for private firms, effectiveness and equity weigh a lot less in the determination of success than efficiency. This means that many different measures are needed to give a clear picture of a public agency's performance. In his classic piece "The Problem of Defining Agency Success", James Q. Wilson (1993) offered an example for policing to point out that obvious measures of success are often illusive: There are no

"real" measures of overall success; what is measurable about the level of public order, safety, and amenity in a given large city can only partially, if at all, be affected by police behavior (Wilson, 1993:160). The point that Wilson is buttressing is that no one measure can encompass success in policing.

Moreover, there are no obvious and natural baselines for many services provided by the public sector. There is no agreed upon point like in the private sector where a firm would have a surplus of revenues that is higher than the marginal gains of the opportunity that was forgone (e.g. investing the money in bonds or the stoke market). When it is possible, like for municipalities in a given country/state/province, success can be relative to what other municipalities are doing. Especially, but not exclusively, for isolated performance initiatives, it is useful to socially determine what is success for a given service. This step has been suggested in the past. The strongest version of this argument has even been that defining what success is has to be done ahead of time if success is to be claimed (O'Donnell and Galat, 2008:100). For one, in their 10-point normative list of how to perform performance management in public health, Moullin and colleagues (2007:284) recommended that managers should "(f)orm a reference group (or steering group) including senior managers, staff, service users and other stakeholders. Identify what a successful service or organisation would look like." In a very similar 10-point normative list of how to perform analytical hierarchy processes for brownfield redevelopment, Wedding and Crawford-Brown (2007:485) advise to "specify the weightings for each indicator of success in some overall ''measure'' of success of an option." The authors go further and add that normatively separating successful from not successful projects that

are already being measured is important; moreover, this assessment should be compared to the opinions of stakeholders (Wedding and Crawford-Brown, 2007:493-494). The commonality in consensus above is that the definition of success should be obtained by implicating all stakeholders, not just the managers. However, there are discordant views on the topic. For example, at the federal level, the Government Accountability Office (Steinhardt, 2008:22) argued that Congress should be the entity defining what success is. This view is not widely shared in the literature. Another minority view asks explicitly that managers should not be the only ones deciding what a success is (Ho, 2008:10).

Since participative meaning on the definition of success is suggested but not implemented, how is success identified by managers performing sensemaking activities? The process is contained and explained at length in the benchmarking, comparisons and target setting section: by using reference points. Reference points are "stimuli of known attributes that act as standards against which other categorically similar stimuli of unknown attributes are compared in order to gain information" (Yockey and Kruml, 2009:97). In plain English, managers identified a performance level as a success by comparing the achieved performance level to other levels of performance. As Moynihan (2008b:104) put it, "analysts are usually struck with interpreting whether performance is satisfactory in light of previous performance or some target, implicit or explicit." As we mentioned before, we also added to the list the performance of other comparable organizations (mean performance, top quartile, etc.) and professional standards as other referent points. The main point here is that a given performance level is labeled and then acted upon, or not acted upon in the case of a success, as a result of comparisons. We will

not reproduce here the lengthy arguments about the benefits, shortcomings and effect of internal benchmarking, external benchmarking, professional standards, and comparisons with targets.

Sensemaking is giving meaning in the lights of reference points, which are comparisons. It suffices here to spell out how these comparisons are specifically operationalized to be equated with a verdict of success. First, perhaps the less labor intensive way to label performance as a success is via internal benchmarking. An improvement of performance becomes a success (Radnor, 2008a:105).

Second, for targets, HM Treasury (2003:12), after extensive consultation with practitioners, defined measurable targets as targets where "there should be a clear and transparent measure of success". Other definitions of success by targets are not as demanding as HM Treasury's. For their part, Marsden and Bonsall (2006:201) urged that "the 'success' of a target should be measured not simply in terms of the achievement of the nominated value of the specified indicator, but in the light of the progress achieved towards fundamental objectives." A less stringent achievement of success through targets is simply meeting the expected performance of preexisting targets (Salminen and Viitala, 2006:24; Behn, 2008:8; Kassel, 2008:242). ""Expected performance," in turn, is defined as achieving, at a minimum, the expected quality of work within a project's anticipated budget and time frame" (Kassel, 2008:242).

Third, there are also those who advocate that since public agencies act for and on the public's behalf, the satisfaction of the public should be the determinant of a success verdict. This made Jamali (2007:381) state that "while internal performance measures may be useful, claims of success cannot be advanced without supporting evidence from concrete outcome measures, of which citizen satisfaction seems the most obvious and critical." Some, like Bolton (2003:23), go as far as to formulate that in the case where so-called objectively measured performance would deteriorate but citizen satisfaction would increase, the overall assessment of that public agency should be deemed a success.

Fourth, another way to ascertain relative performance and eventually success is achieved with quartiles. This method is rather crude but straightforward: the top twenty-five percent of organizations are given as models to emulate (See Higgins, 2005:455-456). This was an element of the Best Value System in Britain (Bowerman, Ball and Francis, 2001:324).

Fifth, an additional way to determine success is via comparisons with some professional standards. One of a few explanations of this phenomenon in the literature comes from Kenis (2006). The author assures us that "professionals do not usually relate to the organization's criteria for success, but more commonly to the criteria that are central to their profession and which they learned during their training" (Kenis, 2006:119).

Sixth, a far less common reference point for interpreting performance and labeling it as a success comes from adjusted performance measurement. Adjusted performance

measurement involves using statistics tools to take into account influences external to agencies in the measurement of performance. It is used by practitioners and academics in some benchmarking systems to offer more comprehensive and fair comparisons, especially in the field of education (e.g. contextual value added in English schools) and health (risk-adjusted performance measurement in NHS hospitals). An organization can be labeled as successful if it surpasses the average performance of other comparable organizations, given the inherent difficulties or advantages that are exogenous to managers (Miller, Kerr and Ritter, 2008:112). This method is more fitting to the public sector than the private sector. The private sector can decide to target a limited market, go after a small niche of clients and close branches that are hurting the bottom line; the public sector cannot.

All in all, the literature tells us that interpreting a given performance level as a success is likely to mean that no corrective or at least strategic actions will be taken. The literature also tells us that, just like sensemaking at large, interpreting performance as a success involves comparisons with reference points. Many such reference points exist. Meeting and/or exceeding those reference points can be interpreted by managers as success.

Failure

From a post-positivist approach, but also from a constructivist approach, albeit from different bases, failure, like success does exist. Evidently, we prefer public services to succeed rather than to fail. We agree with Ellig (2007:7) that flawless organizations, where failures are totally absent, should be looked at suspiciously. This is especially true

for failure as the gap between observed and anticipated performance: given the volatility in which public organizations operate and the external constraints that are placed on them, a flawless public organization is most likely not reporting candidly or not adopting challenging targets (Ellig, 2007:7).

Failure is defined by Meyer and Zucker (1989:22) as sustained low performance. Even though the authors have generic organizations in mind with their definition, it would have different meanings for public and private organizations. As we have mentioned, survival in the private sector stems from economics; survival in the public sector stems from politics.

Failure in the private sector

Failure in the private sector has been defined as "a deterioration in an organization's adaptation to its microniche and the associated reduction of resources within the organization" (Cameron, Sutton and Whetten, 1988:9). Jas and Skelcher (2005:199) explained that in the private sector, failure is understood in similar terms to success: "(…) failure is principally conceived in terms of organizational scale (employees' turnover, product range) or market share (…)". The authors add that these criteria do not translate well for public sector organizations "(…) where scale is a function of legal and political decisions, and services still retain a quasi-monopolistic status" (Jas and Skelcher, 2005:199). Other definitions of failure in the private sector also do not apply to the public sector; like when failure is understood in terms of exit, death, and the mortality of an organization (Mellahi and Wilkinson, 2004:22).

Failure in the public sector

The "collapse and demise of an organization", called macrofailure by Meier and Bohte (2003:105), is rare in the public sector, as organizations can be kept alive indefinitely as they do not have to survive in an economic market, but in a political environment. In the public sector, Meier and Bohte (2003106) tell us that there are instead microfailures: "consistent patterns of poor organizational performance for extended periods of time." A very similar concept of a prolonged poor performing organization is labeled as "permanently failing" by Meyer and Zucker (1989:19). Interestingly, according to the same authors, an organization (in a generic sense) that persists and performs well is not described as succeeding, but as being effective (Meyer and Zucker, 1989:19).

Public organizations have multiple stakeholders. This, coupled to a lack of a bottom line, means that disagreeing perceptions of what can be considered as a success coexist. According to Hall (2007:295) the multiplicity of definitions would be less of a problem for failure than for success: "failures are often so obvious that there is agreement on failure, but lack of agreement on success." Another consequence of the presence of many stakeholders is that in order to take into account multiple view points, many often contradictory objectives will be developed. The measures operationalizing these goals will also be in contradiction with each other. Therefore, a public agency runs the risk of not being able to attend to all of its goals concurrently, which might be perceived as a failure (Van de Walle, 2009:45-46). An illustration of this phenomenon is provided in policing by Rogerson (2006). In England in the 1990s, public services had the tall order to reduce violence, but were also asked to "maintain or increase the number of detections

for violent crimes" (Rogerson, 2006:26). A resulting fall in detections could be seen as a 'failure', the same could be argued for the opposite.

A cautionary remark concerning failure takes roots in the multiplicity of objectives and performance measures. No matter if failure is defined in relation to annual decline, targets, citizen satisfaction, quartiles, professional standards or adjusted performance measurement, an agency, especially a municipality, is unlikely to fail comprehensively (Jas and Skelcher, 2005:199). Therefore, we should reserve the failing label to public services, not public organizations.

As we have tried to make abundantly clear, sensemaking is about comparison to reference points. First, a decline in performance from the previous period is sometimes implied as being a failure. Concurrently, as reported by Geedes and Martin (2000:393), faulting from continuous improvements is sometimes hinted as breeding failure.

Second, for targets, mimicking the argument for success, it is argued that coming short of attaining a target is sometimes defined as a failure (Kaufman, 2006:9). Additionally, the same is also said about not be able to attain all targets (Public Administration Select Committee, 2003:24). This can quickly become problematic for all public organizations, especially for benchmarking organizations with a different number of targets or goals. The Arkansas' version of the No Child Left Behind is a case in point (*see* Miller, Kerr and Ritter, 2008:107). Missing any of the targets translates for schools in Arkansas that they will not meet the *Adequate Yearly Progress* goals. Some of these goals include the

reading and math efficiency of subgroups like White students, African American students, Hispanic students, limited-English-proficient students, low-income students, and special education students. A minimum of forty students is needed for a school to have a recognized subgroup. In effect, like Miller, Kerr and Ritter (2008:107) observed, "schools with more diverse populations are at greater risk of not making AYP because these schools have more opportunities to fail."

Third, rigid classification structures like quartiles and rankings (or. *league tables*, as they are known in the U.K.) have also been used to define failure. The fact that an organization does not rank in the top quartile or at the top of some rankings has been seen at times by the media as a manifestation of failure (Public Administration Select Committee, 2003:24). Independently of the inherent difficulties of performing for public agencies, the arithmetic restrictions on that definition are severe. Strictly speaking, 75 percent of organizations would fail if failure is defined by quartile, and all organizations but one would fail if rankings are involved.

While doing the literature review, we did not find explicit definitions of failure in terms of, citizen satisfaction, professional standards, and adjusted performance measurement. One can wonder if the authors who did write on these referent points in terms of success would agree that the reciprocal of their definitions of success would apply for failure. What we did find in the review of the literature are mentions that failure (but also success) is whatever "higher bodies who bestow legal, financial, or other resources on service providers" say it is (Boyne, 2006:374), according to their judgment. Although this

is nowhere as clear as managerial comparisons to referent points, political verdicts would also be a component of public service sensemaking. However, there is to our knowledge, nothing in the public administration literature to that effect.

There is an argument that has not received coverage in the literature: the nature of public services' failure in the public sector. For once, is failure the reciprocal of success? Is failure the non-occurrence of success, or does it have a distinct, and almost independent identity? Is failure separate from success by a zone of indifference, where performance would instead be thought of as satisfactory and unsatisfactory? A conclusion from the definitions of success and failure above is that there would be some symmetry of definitions in sensemaking. Nonetheless, we can extrapolate from the literature that success and failure are not polar opposites.

Failure over success

The literature on public sector sensemaking, even though it is spotty at times, points in the direction that success and failure are not sides of a same coin. They would not be strictly symmetrical. One of the idea often found in the sensemaking literature in the public sector is that failure would be a dominant label over success. To that effect, Moynihan (2008b:33) tells us that for private firms, the achievement of results is the main focus: the "avoidance of error or malfeasance is of secondary consideration. The opposite is true in the public sector." When it comes to the perceptions of external stakeholders, the perception of one case of failure would be more potent than many successes (Van De Walle and Bouckaert, 2007:1129). Specifically, failures would draw more attention by the media and the citizenry than performance levels deemed successes,

and managers would focus their attention on avoiding failure rather than aiming for success.

The idea that government performance is taken for granted by citizens, but also the media, and that as a result, deficient performance draws more attention has been coined the grievance asymmetry argument by Yang and Holzer (2006:116). Theorizing on the differences between the management in the private sector and in the Canadian federal public sector, Savoie (1999:54) suggested that the avoidance of mistakes and failures might be the main difference between the managing the two sectors.

> It is hardly possible to overstate that control in the public sector is negative and government officials are constantly on the lookout to avoid errors and mistakes, perceived or real. In the private sector, it does not much matter if you get it wrong 30 per cent of the time so long as you can turn a profit at the end of the year and the bottom line remains healthy. In the public sector, it does not much matter if you get it right 95 per cent of the time because the focus will be on the 5 per cent of the time you get it wrong. (Savoie, 1999:54)

Undue attention to failures, while successes go unnoticed (Hall, 2007:285), would feed the "unfair perception that our public agencies are poorly run and that all government authorities are venal" (Balk, 1993:439).

Describing the environment in which local authorities would evolve in the U.K., Davies (2004:43) commented that there is also the propensity in the media to focus on the "sensational rather the mundane." The author goes on saying that "(…) an excellent rating is unlikely to hit the headlines while the poor performers will usually get plenty of publicity" (Davies, 2004:43). One of the few empirical studies of the electoral impact of incumbent local administration also suggests that for electors, failure carries more weight

than success. James and John (2006) analyze the support for the incumbent municipal government between 1999 and 2003, controlling for possible influences like performance variation over time, variation in the local tax level, variation in local economic circumstances, and if the incumbent local party is affiliated with the incumbent party at the national level. What sticks out of the authors' findings is that the overall performance label (*either* poor, weak, fair, good, or excellent) is influential on voting patterns, whereas increasing performance trends are not (James and John, 2006:575). Local authorities with a CPA score of 'poor' lost 6% electoral support compared with local authorities with 'weak', 'fair' or 'good' scores (James and John, 2006:574). The findings make the authors state that "while the poor CPA has a clear impact, the excellent CPA category is not significant, suggesting that poor performance is punished but equally strong (in terms of absolute difference in level of category from the middle fair category) excellent performance is not rewarded" (James and John, 2006:575). To offer a tentative explanation about the negativity bias of electors, the authors turn to sensemaking, more specifically to the shape of notions of failure and success. What they say about sensemaking is that:

> (…) comparative, yardstick, performance may be an important influence on individuals' assessments of local authorities. The poor label may have greater salience if it is seen as being further below some notion of average (perhaps mean or median) local authority performance than excellent is above such an average. (James and John, 2006:578)

In essence, James and John opine that failure and success, at least for electors, are not symmetrical polar opposites. For public sector external stakeholders, from a starting neutral zone, the intensity or distance of a failure is greater than the one for a success.

Managers would be more preoccupied by steering clear of failures than by trying to achieve successes. In the absence of articulated performance measures, staying off the front page of the local paper could be seen as a proxy for performance by many municipal managers (Aaron, 2008:28). According to Try and Radnor (2008:668), behaviors like selective reporting would be a direct consequence of the risk adverse culture present in the public sector. Publicizing poorly performing programs would have limited appeal in an environment where there is "limited rewards for success but considerable punishment for failure" (Try and Radnor, 2003:668-669). Reverberating on the same reality, Plant and Douglas (2006:43) add that this emphasis on failure over success is a feature unique to public organizations. The most insightful and complete analysis of the dominance of failure avoidance over success achievement in the public sector was put together by Johnsen (2008), when concluding on rankings and performance assessment of public education in Norway. The essence of his analysis is worth reproducing here.

> (…) low performance gets most of the attention and this in spite of the "learning from best practice" rhetoric. In public management, as opposed to business management, there is a tendency to "learn from bad performance." In the public sector, avoiding low performance that could result in "naming and shaming" could be as strong an incentive as performing well. There might not be any strong an incentive as performing "best" because the "winner" hardly "takes it all." For many public services avoiding low performance, by achieving a certain basic or average level of performance for specific (often vulnerable) users and clients, could be more important than achieving a high level of service. Therefore, in public policy and management, it may be more important to avoid being "bad" than being "best". (Johnsen, 2008:172)

Johnsen's description of (municipal) public education in Norway covers the main idea of failure avoidance in the public sector.

The main argumentation in this section is that external stakeholders pay more attention to failures than successes. There would be no payout, at least in public recognition, for excellence in public services. Perhaps as a result, public managers tend to channel their energy in failure avoidance instead of achieving successes. One point that we try to make clear, is that there is little empirical evidence supporting the widespread view that failure dominates over success in the public sector. Most of the literature that was presented is speculative.

In terms of sensemaking, one more issue has to be covered before moving forward with this study: is there something that exists between success and failure?

Satisficing

The idea of satisficing is credited to Herbert Simon (1945), in Administrative Behavior. Simon's idea was to offer an alternative model to the rational economic model of maximization, which he perceived as unrealistic regarding the cognitive and analytical abilities required by *homo economicus*. Despite the fact that Simon's idea of bounded rationality was awarded a Nobel Prize in economics, other ideas linked to it like *satisficing*, had relatively low traction in public administration empirical research. Many reasons could be advanced to explain why a powerful idea like satisficing received little scholarly attention. Speculatively, the relative strength of economics as a social science discipline and the influence of economics' concepts and theories on other disciplines could be mentioned. Another explanation is that the mathematics behind maximization is

rather straightforward, whereas satisficing is fuzzier and much more difficult to formulate as an equation.

All in all, the express mention of satisficing in public administration scholarly and practical work is very rare. Even the assumption about bureaucratic behavior is often borrowed wholesale from economics, where bureaucrats are hypothesized as budget-maximizers (Niskanen, 1971) and rational principled-agents taking advantage of information asymmetry. Public agencies become populated by managers described in such ways that they would be interchangeable with their counterparts in the private sector. For our purposes in this research, we took our distances from the debate about the applicability of economic and private sector theoretical assumption and looked pragmatically at what the literature had to say about the use and sensemaking of performance information by public managers.

The main discovery is that the literature in performance measurement says very little about the nature of public managers in their performance management activities. Even the mention of satisficing in research design occurs on the rarest of occasions (*for the only such example we found, see* Yockey and Kruml (2009:103)). As for a singular real world inclusion of satisficing into a performance measurement initiative, the *Productivity Measurement and Enhancement System* (ProMES) in policing in Sweden included satisficing provisions, where minimal standards on performance indicators were achieved through participation (Pritchard *et al.,* 2009:88). Later performance assessment and sensemaking was done form the basis of comparisons of what was considered minimally

satisfying performance levels (Pritchard *et al.*, 2009:88). In order to witness the inclusion of the concept of satisficing, one has to turn his/her attention to the academic literature, and keep a very open mind to what could constitute satisficing behavior in performance measurement.

While theorizing on the maximizing/satisficing behavior of private and public organizations in benchmarking, Bowerman and colleagues (2002) suggested that private organizations would tend to engage in world-class benchmarking, comparing across different industries to learn from the best and become the best at certain activities. Basing their arguments on Spendolini (1992: 113) and Cox and Thompson (1998: 3), Bowerman and colleagues (2002:433) expressed that for businesses, best practices, that is a maximizing attitude to performance management, are the prime foci. In contrast, the authors suggest that public organizations use benchmarking to show that they are good enough or at least not the worst, rather than try to prove that they are the best. (Bowerman *et al.*, 2002:433). The empirical findings of Bowerman and colleagues (2002) from 80 interviews and 725 questionnaires of local government managers on benchmarking support their previous assertions. Frequent comments from their interviews were that “"I suspect we will be well ahead". In other words it gave comfort that they were "good enough" and gave them ideas about how to "become better" rather than indicating ways to "be the best"” (Bowerman *et al.*, 2002:433). From the results of that study, one could cautiously advance that benchmarking in the public sector is not about maximizing. A critical reader would have to read again the previous sections on the *types of benchmarking*, *use of benchmarking* and on *targets* to make up his/her mind

regarding the performance measurement behavior of public sector organizations. At present, before the results of the survey on what do managers in Quebec use as bases of comparisons and bases for target setting, we would suggest that public managers are following *satisficing* rather than *maximizing* principles in performance measurement. This is to say that managers, in their sensemaking activities, are less likely to think in terms of 'success' and 'failure' and more likely to think in terms of 'satisfying' and 'unsatisfying'.

The review of the literature begs the following question:

> **R4. While interpreting their own performance, when do municipal managers deem that the municipality's performance is (a) a failure, (b) unsatisfying, (c) satisfying, (d) a success?**
>
> **H11. It is expected that for sensemaking purposes, managers will reflect more in terms of satisficing than in terms of maximizing.**
>
> **H12. It is expected that for sensemaking purposes, managers' verdict of performance will lean more in positive terms of satisfying and success than in negative terms of unsatisfying and failure.**
>
> **H13. It is expected that for sensemaking purposes, managers will be more likely to interpret given performance levels in ways that put their performance in the best possible light.**

Chapter 2: Research setting[††]

The present research is set in the province of Quebec, Canada. The role and characteristics of municipalities in Canada are almost identical to that of municipalities in the United States. The only important difference is that Canadian Municipalities are not responsible for education. Education is a provincial responsibility. Readers accustomed to the workings of municipal administrations in the United States can assume that what they know about municipalities also applies in Canada

The *Municipal Management Indicators* system in Quebec

Political background

Quebec's municipal performance measurement system came later than other municipal performance measurement systems in Ontario, Canada, the U.K. and New South Wales, Australia. This delayed implementation could be credited in part for leapfrogging pitfalls that were encountered in previous performance measurement systems: hastiness of implementation (Chang and Kelly, 1994:13), no or little practitioners outreach in the design phase (Davis, 1998), lack of comparison subcategories (Foltin, 1999:44), and absence of shared accounting practices (Coe, 1999:114). The system was tried in pilot projects between May 2001 and May 2002. Since 2005, it is mandatory for municipalities to make the data public. According to article 17.6.1 of the *Law on the Ministry of Municipal Affairs and the Metropolis* (L.R.Q., C. M-22.1), under article 241 of chapter 37 of the laws of 2002, the Ministry of Municipal Affairs, Regions and Territorial Use

[††] This chapter appears in *Schatteman, A.M. and E. Charbonneau (forthcoming). "A Comparative Study of Municipal Performance Measurement Systems in Ontario and Quebec, Canada," International Journal of Public Sector Performance Management 2(3)*

values for the mandatory indicators of the Municipal Management Indicators (CPEGM, 2005a:38).

Québec's municipal performance measurement system survived a regime change in 2003, when a *Parti Québécois* government was replaced with the Liberal Party of Québec. It also thrived under two *Parti Québécois* ministers and two *Liberal Party* ministers. The Québec Ministry of Municipal Affairs, Regions and Territorial Use (MAMROT in French) implemented the *Municipal Management Indicators* in 2003. Québec's Municipal Management Indicators originates from consultations in 1999, when stakeholders from six professionals associations, two associations of municipalities, one Business school and Québec's Ministry of Municipal Affairs and Regions[‡‡] met to discuss the plan of Québec's municipal performance measurement system.

Performance measurement system components

The *Indicateurs de gestion municipaux* stands on its own. Accordingly with Québec tradition of deliberation, the Municipal Management Indicators performance measurement system is not tied with financial consequences. The Ministry of Municipal Affairs, Regions and Territorial Use outsourced the analysis of the data to a third party, the *Center for the Promotion of Municipal Management Excellency* (CPEGM in French) at the *École des Hautes Études Commerciales de Montréal* (HEC-Montréal). The

[‡‡] Despite its French heritage, the government of Québec operates like the other Canadian provinces, in a Westminster-style parliamentary system. One of the consequences is that the portfolios of agencies are subject to change. For example, the Ministry of Municipal Affairs has also been responsible in time for the Metropolis [Montreal], Sports and Leisure, and Regions. It is currently known as the *Ministère des Affaires municipales, Régions et Occupation du territoire*, which loosely translates into Ministry of Municipal Affairs, Regions and Territorial Use.

CEPGM is the depository of the data for the Ministry of Municipal Affairs and Regions. The CPEGM produces a report where municipal managers can compare their performance with aggregated data of municipalities of their size. The data in the reports are presented by quartiles, by population sizes. Although 'best practices' are supposed to be offered to municipal managers, it never materialized in reality. The closest thing to best practices that is offered to municipalities is an anonymous upper limit on a box plot chart that comes in reports: specific examples of municipalities to emulate are not yet offered.

Performance measurement requirements

In Québec, for the 2003-2006 period, the *Municipal Management Indicators* required municipalities to collect information for 19 mandatory indicators. Of those 19 indicators, there were three indicators about street maintenance, two about snow removal, five on water treatment and distribution, two about sewage systems and seven about global financial health. Eight out the twelve non-financial indicators were cost indicators. In 2007, changes were made in terms of what is collected in terms of indicators. Some indicators that were deemed problematic from municipalities and are no longer required by the Ministry of Municipal Affairs, Regions and Territorial Use. There is now one indicator about street maintenance, one for snow removal, four on water treatment and distribution, two about sewage systems and two about global financial health. In addition, four new indicators on human resources have been added. Nine out of the eleven non-financial indicators measure cost. Moreover, thirteen indicators are now facultative from 2007 onward: two of these indicators measures dimensions of human resources, three

facultative indicators are on fire services, five are about culture and leisure and finally, three are have to do refuse collection & recycling. These facultative indicators will possibly become mandatory in a few years.

Contextualization is a recurrent theme in all official reports on the *Municipal Management Indicators*. It is specified in every official document by the Ministry that "(...) the interpretation of values obtained for the indicators will often be different between municipalities, depending on realities specific to municipalities and the service at hand[§§]" (Ministère des Affaires municipales, du Sport et du Loisir, 2004:4). Contextualization is supposed to offer "(…) reasonable expectations and potential improvements" (CPEGM, 2005a:10). Originally, the indicators were developed with two aims in mind: to help elected officials and managers to improve the management of municipal services and report to citizens (Guindon and Bellavance, 2004:3; CPEGM, 2005a:8). It is specified in the guide sent to all municipalities at the beginning of the implementation phase that "Any external comparisons makes sense only if influential factors are known for each municipalities included in the comparison" (MAMSL, 2004:5).

Performance reporting requirements

Initially, the provincial government of Québec has made some efforts to facilitate reporting to citizens. First, the SESAMM software has a built-in reporting function that makes reporting of raw data straightforward. Second, the provincial government produced a guide to municipalities about reporting practices. In this guide, practical tips

[§§] Translation by the author.

are given, along three comprehensive examples of reports that should be emulated (CPEGMa, 2005, Annex A, B, C). However, the current objectives of the system no longer include elements of citizen reporting.

The provincial guidebook to municipalities stipulate that comparisons with other municipalities should be sought (MAMSL, 2004:9), mainly for voluntary performance improvement purposes, and reporting purposes. However, so far municipalities transmit data online unilaterally, through the SESAMM software, to the provincial government (MAMSL, 2004:8) but they do not have access to current pooled municipal data in real time. This means that they cannot readily use current external information to set performance targets or report information taking full advantage of the performance measurement system to contrast their performance with best practices. However, municipal managers have access to two source of information to contextualize their performance. First, they can compare their achieved performance in regards of the rest of municipalities in the province but only through statistical aggregates, through the use of quartiles. A report containing performance organized by quartiles is published with a one to two-year delay by the Centre for the Promotion of Excellence in Municipal Management at HEC Montréal, a prominent business school who contracts with the provincial government on the municipal performance measurement system. Second, municipal managers now have a tool inspired from Ontario MIDAS system. They can consult comparative performance with a password-protected online tool. However, the data from this tool accused a two year delay and are aggregated and anonymous. Third, a list of predetermined influential factors is provided by the government of Québec to help

municipalities contextualized why they perform the way they did (Ministère des Affaires Muncipales et Régions, 2005).

Municipalities have to report on the mandatory indicators once a year to the provincial government. Municipalities must transmit the information by September 30 each year (MAMSL, 2004:4; MAMR, 2008). For example, it means that municipalities had up to September 30th of 2009 to report on 2008 data. This largely explains why the report with performance quartile comes in so late. Reporting on the performance indicator is mandatory; reporting on influential factors is not (MAMSL, 2008).

Chapter 3: Methodology and Data

The present research utilizes the case of municipalities in Quebec to study the use and sensemaking of performance information by managers. Since 2003, all municipalities in that Canadian province are mandated to collect and report a set of standardized performance indicators. The methodological approach for this research is twofold. First, an online survey was sent to all General Managers in the 1113 municipalities in Quebec in the winter of 2009-2010. Perception data from that survey are coupled with so called hard performance information on the values of the performance indicators. The survey also had two open-ended questions that will supply data for the qualitative chapter of the dissertation. Second, the Partners on Municipal Management Indicators Committee [*Comité des partenaires des indicateurs de gestion municipaux*], an umbrella organization of professional associations and governmental agencies, organized a full day of activities on January 28th, 2010. One of the activities consisted of focus groups about selected findings in the survey mentioned earlier. The results from the focus groups supply the other part of the qualitative component of this research.

Method #1: Online Survey

One of the methodological tools used in this research consists of an online survey. Since the aim of the research is to uncover the analytical use of performance information that is sensemaking, a survey is an adequate tool to fulfill the mandate of the present research. An electronic survey was sent to public managers in every municipality in the province of Quebec. What motivates this choice is that official correspondence from MAMROT is sent electronically; MAMROT routinely contacts municipalities by electronic means and

sets-up performance transition tools that are computer-based. The contact list of General Managers provided by MAMROT consisted of electronic addresses. The sheer number of municipalities in the province of Quebec involves prohibitive costs for mail surveys. It also has other methodological implications: the large number of municipalities is a serious hindrance to a large number of face to face interviews. Additionally, the distances between municipalities are great in Quebec, which is the vastest Canadian province. An e-mail survey seemed like the best compromise between an online survey and a mail survey.

An identification code on the surveys made possible to join the values for the performance information from MAMROT's dataset, the values obtained for the different dependent variables, and the independent variables from the survey. Juxtaposing "hard" and survey data has been done in recent landmark performance measurement research on Norwegian local politicians' attitudes towards comparative evaluation of local bureaus' performances against other jurisdictions (Revelli and Tovmo, 2007). It has been argued that "progress in building a science of public management has been slowed by the lack of widely available data sets that permits a complete specification of the factors involved in determining organizational performance" (Boyne *et al.*, 2006:303). When studying performance information use and sensemaking, joining performance data and survey data would further the field understanding, since:

> Government data sources often provide a rich set of performance indicators but generally contain little or nothing that directly taps management or could be interpreted as a measure of management. (…) Archival or government data sets need to be merged with survey data sets that add management, structure and other variables into the mix. (Boyne *et al.*, 2006:303)

Electronic surveys have the advantage to offer an option to track responders, without offering an overt reminder that their answering patterns can be traced back to their individual municipalities.

Electronic surveys were not used in combination with mail survey: only one survey format was used. The motivation behind this methodological choice comes from Kampen's (2007) study of government performance. In this methodological study, the author compares the response patterns of mail and face-to-face surveys. The results are that different survey format introduced undue biases that threatened the internal validity of the study (Kampen, 2007:810-811).

The survey instrument (Appendix) included a total of nine series of closed-ended questions, and two open-ended questions. General Managers were asked to qualify their general use of performance information; their management, financial and reporting use of performance information; and their comparative use of performance information. The two open-ended questions were present in the survey. The first of these questions was targeted to General Managers who indicated to question1 that they did not use performance indicators at all after they are collected and transmitted to MAMROT. The second group of open-ended questions were available to all survey-takers. It was phrased "If you have comments to formulate regarding the management indicators or the present survey, please add them in the space below." Additionally, and unlike the different uses mentioned earlier, the question about the sensemaking use was scenario-based and more

hypothetical. The use of fictional scenario for sensemaking research has at least one precedent (Wagner and Gooding, 1997: 280).

The survey instrument was pretested with twelve active members in the Partners on Municipal Management Indicators Committee in the Fall of 2009. Many of the Committee members who were involved in the pretesting of the survey are current or former municipal General Managers and/or Chief Financial Officer. The final version of the survey was reviewed and approved by two ethical supervisory bodies: *HEC Montréal Ethical Board* and *Rutgers University Institutional Review Board* (IRB) for human subject protection in research.

Population and Sampling

The values for the different dependent variables and the independent variables originate from survey data. The survey was sent to all 1113 municipalities in Quebec. Surveying the whole population instead of a sample, given a hypothetical fixed response rate, provides this study with more observations. This higher *n* from a population rather than a sample would add to the external validity of the descriptive statistics of the study, if the sample is representative of the population. Moreover, a higher *n* brings about more statistical power. That is to say that all other factors remaining constant, a higher *n* permits uncovering correlations that otherwise would have been below agreed-upon thresholds. Similarly, a higher *n* makes it possible to construct a more complex model using more independent variables.

The e-mail addresses for General Managers were transmitted by the *Ministry of Municipal Affairs, Regions and Territorial Use*. The data of this study uses the whole population of municipalities in Quebec. The fact that the MAMROT had electronic contacts of every municipality dissipates some of the concerns about a digital divide that could be potentially present in the province of Quebec, given the proportion of municipalities of less than 5,000 residents. Table 6 summarizes the demographics of these municipalities.

Table 6: Quebec Municipalities by Population, in 2008, including Survey Participation

Size of Municipalities	**Number of Municipalities**	**Survey Participation**	
		No	**Yes** (% Yes)
0 to 499	206	122	84 (41%)
500 to 999	272	174	98 (36%)
1,000 to 1,999	261	179	82 (31%)
2,000 to 2,999	114	79	35 (31%)
3,000 to 4,999	91	61	30 (33%)
5,000 to 9,999	73	51	22 (30%)
10,000 to 24,999	55	37	18 (33%)
25,000 to 49,999	23	11	12 (52%)
50,000 to 99,999	9	4	5 (56%)
100,000+	9	5	4 (44%)
Total	**1113**	**723**	**390** (35%)

Source: Adapted from Ministère des Affaires municipales et Régions (2008)

As we observe in table 6, 390 municipalities out of 1113 replied to the survey, for a response rate of 35 percent. What contributed to the 35-percent response rate were population-stratified reminders done by the researchers, but also by professional associations active in the Partners on Municipal Management Indicators Committee, like the *Corporation des officiers municipaux agréés du Québec* (COMAQ). Interestingly, the response rate was rather consistent amongst municipalities of different sizes. However, since we are interested in the perception of General Managers, we had to lay aside surveys filled out by other managers. The number of suitable surveys for the quantitative analyses for this research is 312.

Unit of Analysis

The unit of analysis for this study is the General Manager, the highest-ranking municipal administrative employee. The independent variables on the use and especially sensemaking, are perception-based measures. The perceptions of the top manager are what the research is based on.

It has been recognized in past research on use of performance measurement information that surveying a single manager per organization constitutes a limitation (Lægreid, Roness et Rubecksen, 2008:45). Municipal managers working in finances and budgeting do not report their activities according to the same criteria as other managers (Marcuccio et Steccolini, 2009 :160). A significant difference exists as well regarding perceived problems of performance measurement for finance and budgeting managers, and other managers (Willoughby, 2004:36).

The intention was to send the survey to two public servants per municipality: the *trésorier* [Chief Financial Officer] and the *directeur général* [General Manager]. There are two reasons that motivated that choice. First, reaching out to more than one official increases the chance to have at least one of them fill the survey and provide survey data for their municipality. Second, the Chief Financial Officer and the General Manager of a given municipality might diverge in their use and interpretation of performance information. Having the possibility to access two points of views for each municipality would render a more complex picture of performance measurement activities.

Aware of the perils just mentioned, Andrews, Boyne and Enticott (2006 : 282-283) studied 120 English local authorities in regard of the CPA benchmarking system, by surveying 1257 managers : that is more than ten managers per municipalities. Accordingly, a modest scaled-down version of Andrews, Boyne and Enticott's (2006) sampling strategy was included in a formal request sent to the *Ministry of Municipal Affairs, Regions and Territorial Use* of Quebec to survey more than one manager per municipality. In the proposed plan, because 66% of municipalities had less than 2,000 residents in 2008, only municipalities with more than 2,000 residents, 380 of the 1113 municipalities in Quebec, would be sent two surveys. The intent was to take into account that for many of these small municipalities, the Chief Financial Officer and the General Manager are the same individual. This request to survey more than one public servant by municipality was denied by the *Ministry of Municipal Affairs, Regions and Territorial Use*. Therefore, amidst the downsides, only one survey was sent in each municipality.

Data #1

The data used in the quantitative part of this research come from the self-administered survey described earlier on the behaviors and perception of managers toward the management indicators, and the values of the mandatory indicators. Some basic socio-economic characteristics of the municipality where managers operate are also included. Many survey instruments that were used in the survey of Quebec's General Managers come from previous studies about managerial behavior related to performance measurement and management. However, the survey instrument developed for the realities and specificities of performance measurement regimes of other countries were adapted to fit Quebec's municipal benchmarking system reality.

Dependent variable: use of performance information

Knowing the proportion of local government managers who disclaim that they are using performance information is appealing in its own right. After all, resources and efforts are deployed by MAMROT and municipal mangers to collect, compile and analyze the information. As an information-based tool, performance measurement needs to be used if its benefits are to be felt. Additionally, use is a necessary condition for sensemaking.

Three different kinds of uses can be derived from the survey: general use, specific use (management, budgeting, reporting), and comparative use. Table 7 offers a glance on how the questions are constructed.

Table 7. Dependent Variables related to Uses of Performance Information

Dependent Variables	Survey Question	Citation
General use	Question 1 Data collection and reporting municipal management indicators has been mandatory for all municipalities since 2003. According to your observations, what is the utilization level in your municipality?	Developed for this study
Use for management, budgeting and reporting	Questions 2-3 From what you have observed, indicate what are the reasons for which management indicators are used in your municipality. Indicate if since their implementation, mandatory management indicators for the different functions and activities have been explicitly mentioned [appeared] in the preparation of the budget and the annual report on the financial situation in your municipality.	Adapted from Rogers, 2006
Comparative use	Question 4 For each mandatory indicators presented in the table below, first indicate if you ever compared the value of the indicator, and second, if you ever established targets for this indicator. If you checked "yes" for either one of these questions, elaborate by checking all boxes that applies in the table.	Similar to Taylor, 2009

General use

The first question of the survey asks managers how often are the municipal management indicators used in their municipality. This question differentiates between managers who do not use performance information at all and those who do even occasionally. Managers who do not use management indicators passed over the question on more specific uses of performance information. The first question does not differentiate between *symbolic* and *actual* use of performance information. It is akin to the first generation studies on the use of performance indicators.

The operationalization of this variable is straightforward: for managers that expressed that the municipal management indicators "are not used", the variable was coded "0"; for the others that expressed that they were at least seldom used were coded "1".

Table 8. Descriptive Statistics of Dependent Variables – General Use

Data collection and reporting municipal management indicators has been mandatory for all municipalities since 2003. According to your observations, what is the utilization level in your municipality? *(1=at least seldom used)*	**Question**	**n**	**min**	**max**	**mean**	**SD**
General use	1	312	0	1	0.446	0.498

Specific use

A number of existent survey instruments measuring use of performance information were presented in the literature review. Roger (2006) differentiates between different uses like, use for management, use for budgeting and use for reporting. This differentiation of uses makes this GASB instrument more precise than other survey instruments measuring managerial use of performance information. For this reason, the GASB instrument was adapted to the reality of Quebec's municipal environment.

In her dissertation, Rogers (2006) regrouped survey questions on the use of indicators for three different functions. In her models, there were seven items for management uses, six items for budgeting uses, and three items for reporting uses. For all the different questions items, respondents were asked offered five options about the proportion of

departments using them, from no department at all to all departments. An additive index was set up for each use. The management use index spanned from 7 to 35 points, the budgeting use spanned from 6 to 30, and the reporting use spanned from 3 to 15. The statistical internal reliability scores for Cronbach alphas were respectively of .956, .934, and .898.

To keep true to the realities of Quebec's benchmarking system and after demands from the twelve early survey takers, the number of items for each different uses were reduced. In the current survey, there are four items for management uses, three items for budgeting uses, and three items for reporting uses. Also, to reflect the fact that the survey is sent to many small and very small municipalities and not to municipalities of more than 25,000 residents like GASB's sample, the survey takers only had to express if items for specific uses were indeed used. For municipalities in Quebec, because the majority of them are small, it did not make sense to ask the proportion of departments that were using indicators for specific functions. The mean for the items constituting management use, budgeting use and reporting use are all low. These uses are not widespread in municipalities in Quebec. For the three specific use indices, the coding is dichotomous. The indices are coded '1' in the occurrence of at least one of the reported uses, and '0' if it is not the case.

Table 9. Descriptive Statistics of Dependent Variables – Management Use

From what you have observed, indicate what are the reasons for which management indicators are used in your municipality *(1=observed)*							
	Question		**n**	**min**	**max**	**mean**	**SD**
Management use	2		312	0	1	0.201	0.402
		Statements					
	2.2	In establishing contracts for services	312	0	1	0.019	0.138
	2.3	Managing operations or routine decisions	312	0	1	0.042	0.2
	2.4	Evaluation to establish underlying reasons for results	312	0	1	0.141	0.349
	2.5	Specific performance improvement initiatives	312	0	1	0.08	0.272

The same logic applies to the indices for budgeting use and reporting use of performance information.

Table 10. Descriptive Statistics of Dependent Variables – Budgeting Use

From what you have observed, indicate what are the reasons for which <u>management indicators are used in your municipality</u> *(1=observed)* **& presence in Annual budget or Annual report on the financial situation**							
	Questions		**n**	**min**	**max**	**mean**	**SD**
Budgeting use	2&3		312	0	1	0.330	0.471
		Statement					
	2.1	To prepare budgets, including resources allocations or discussion of resources <u>reallocations</u>	312	0	1	0.083	0.277
	3.1	<u>Explicitly mentioned in budget</u>	312	0	1	0.231	0.422
	3.2	Explicitly mentioned in financial report	312	0	1	0.215	0.411

Table 11. Descriptive Statistics of Dependent Variables – Reporting Use

From what you have observed, indicate what are the reasons for which management indicators are used in your municipality *(1=observed)*							
	Questions		**n**	**min**	**max**	**mean**	**SD**
Reporting use	2		312	0	1	0.362	0.481
		Statement					
	2.6	To provide feedback to managers and <u>employees</u>	312	0	1	0.090	0.286
	2.7	<u>To report to elected officials</u>	312	0	1	0.272	0.446
	2.8	To report to citizens, citizen groups or to inform the medias	312	0	1	0.147	0.355

The idea of additive indices was not kept. The statistical internal reliability scores for Kuder-Richardson coefficient of reliability for the management use, budgeting use and reporting use are respectively of of .49, .61, and .49. For Roger's (2006) indices, a higher value of an index reflected a more intense use of performance information for management, budgeting and reporting. In the current research, any manifestation of management, budgeting and reporting use is considered as use. Many factors can influence the use of performance information. Table 12 summarizes the factors that are accounted for in this research.

Table 12. Independent Variables related to Barriers to Use, Internal Characteristics of Municipalities, and Performance

Independent Variables	Survey Question	Citation
Barriers to Performance Measurement Use	Question 7 Some managers identified barriers which would limit the use of management indicators in decision making. Indicate your level of agreement regarding the following statements on the management indicators.	Adapted from Siverbo and Johansson, 2006
Internal characteristics of the municipality	Question 8 Indicate your level of agreement regarding the following statements on the management indicators, *on the current situation in your municipality.*	Adapted from Boyne and Enticott, 2004; Andrews, Boyne and Enticott, 2006
Performance	Values of the indicators for 2006-2008 period	MAMROT's official data
Size of budget	Log. of total revenue in 2006	MAMROT's official data
Size of population	Population in 2008	MAMROT's official data
Size of municipality	Area in km^2	MAMROT's official data
Administrative region	Administrative region of the municipality	MAMROT's official data

Independent variable: barriers to uses of performance information

In the literate review section on the use of performance measurement, it was made clear that many previous explanatory models of performance management use were conducted in the United States. This impacted the independent variables that are included in the different models. Since the municipal benchmarking systems in the United States are voluntary, and not mandatory and systematic, perceived benefits are included in statistical models to explain use. This makes intuitive sense. If a municipality decides to join a network or start up a performance measurement initiative, it is in part because benefits from performance management are expected. For example, the study from which the dependent variables were adapted, that is Rogers (2006) study using GASB's survey data, went down this route.

In a context of a mandatory and systematic benchmarking system, a level of government imposes on another level of government the participation in a performance measurement project. Municipalities are bounded and constrained, at the very least, to collect and transmit information on standardized performance indicators. In such a case, although it is not impossible that the managers who are to go further and use performance information, it would be less because of perceived benefits, but rather because they do not encounter barriers. After all, municipal managers are forced to minimally participate, and are expected to use the information. The level of government mandating the collection and transmission often define the benefits that are expected out of the benchmarking system. It was the case of Quebec's benchmarking system from the

beginning (MAMSL, 2004). Because of the mandatory nature of the system, perceived benefits were dropped in favor of perceived barriers.

Siverbo and Johansson (2006:283-284), in their study of the voluntary RPE municipal system in Sweden, used a survey instrument to measure the perceived barriers to performance measurement implementation and use. Their survey instrument was adopted, and then adapted to fit more closely to the reality of municipalities in Quebec. Similarly to almost all previous studies of performance measurement use, Siverbo and Johansson (2006) offered descriptive analyses and no regression analysis. This limited the directives that could have been derived from a previous operationalization of this variable.

The researchers established three categories of barriers of use: those related with being unwilling to use performance information, those related to being unable, and those related to being prevented from using performance information. Each barrier was constituted of four items. Again, because regressions were not run in the Swedish study, there are no score of internal reliability to be assessed. In the current survey, a total of fourteen items are present. At the demand of the early survey takers and MAMROT, one item, measuring being prevented from using performance information, was dropped. The item was "the municipality has an explicit or implicit policy against the municipal management indicators." The rest of the eleven items from Siverbo and Johansson's (2006:283-284) instrument were barely altered. Two more items were added to the list. In the survey, four items constitute the 'unwilling' portion of barriers, six items constitute

the ‘unable’ portion, and three items constitute the ‘prevented’ portion. Additive indices akin to the one Rogers’ (2006) used for the dependent variables were developed. On every item, surveyed managers could identify if the statements described the reality in their municipality by expressing if they ‘agree’, ‘somewhat agree’, ‘somewhat disagree’, or ‘disagree’ with the statement. A full disagreement with the perception that the barrier applied in their municipality was code ‘4’; a full agreement was coded ‘1’. Therefore, the unwillingness barrier spans from four to sixteen points, the inability barrier spans from six to twenty-four, and the prevented barrier also spans from three to twelve.

Table 13. Descriptive Statistics of Independent Variables– Unwillingness Barrier

Some managers identified barriers which would limit the use of management indicators in decision making. Indicate your level of agreement regarding the following statements on the management indicators.
(1=disagree to 4=agree)

	Question		n	min	max	mean	SD
Unwillingness barrier	7		286	4	16	12.259	2.86
		Statements					
	7.1	Management indicators are not considered useful	296	1	4	3.166	.873
	7.2	Management indicators are not trustworthy	291	1	4	2.670	.969
	7.3	Management indicators are felt to convey an incomplete picture of the organization	293	1	4	3.113	.874
	7.4	We fear that management indicators are misunderstood and misinterpreted	295	1	4	3.366	.757

Table 14. Descriptive Statistics of Independent Variables – Prevented Barrier

Some managers identified barriers which would limit the use of management indicators in decision making. Indicate your level of agreement regarding the following statements on the management indicators. *(1=disagree to 4=agree)*							
	Question		**n**	**min**	**max**	**mean**	**SD**
Prevented barrier	7		276	3	12	7.783	2.151
		Statements					
	7.11	Our officials are uninterested in the management indicators	288	1	4	3.285	.836
	7.12	Management indicators are seen as a threat	288	1	4	2.229	.964
	7.13	Management indicators will expose our weaknesses	286	1	4	2.259	.942

Table 15. Descriptive Statistics of Independent Variables– Inability Barrier

Some managers identified barriers which would limit the use of management indicators in decision making. Indicate your level of agreement regarding the following statements on the management indicators.
(1=disagree to 4=agree)

	Question		n	min	max	mean	SD
Inability barrier	7		270	6	24	18.556	4.176
		Statements					
	7.5	We do not know how to integrate management indicators into decision making	291	1	4	2.948	.921
	7.6	We are not able to access data that would enable us to compare our results to similar municipalities	288	1	4	3.045	.956
	7.7	We lack the time to use management indicators	292	1	4	3.364	.849
	7.8	We lack the staff with the expertise to work with management indicators	294	1	4	3.251	.962
	7.9	We lack the computerized tools to gather the detailed data on the management indicators	287	1	4	2.944	.940
	7.10	We need additional information to use the management indicators	285	1	4	2.993	.979

The statistical internal reliability scores for Cronbach alphas for the three indices are respectively of .835, .842, and .676. A higher value of an index reflects more numerous and intense perceived barriers for the use of performance information.

Independent variable: internal characteristics of the municipality

Managerial practices can help explain the use of performance information. A survey instrument suited to study performance management was needed. Such an instrument exists. Initially developed to understand the characteristics of local authorities in the U.K. under the Best Value System (Enticott *et al.*, 2002), the instrument was later used by the Audit Commission and academics. Academic researchers sampled the lengthy survey to circumscribe the instrument to five dimensions that have been identified by the Audit Commission (2002:3-4, *in Boyne and Enticott, 2004:12*) has differences between excellent, good, fair, weak and poor local authorities. In their study of local authorities' performance Boyne and Enticott (2004) used twenty-five questions from the instrument to measure internal characteristics of local authorities. The twenty-five items were related to the five dimensions identified by the Audit Commission. There were six items for 'effective leadership', five items for 'management arrangements', three item for 'clarity of priorities', five items for the 'links between priorities and community needs', and six item for 'cross-cutting/partnership working' (Boyne and Enticott, 2004:14). No values for the internal validity of the indices were offered.

From the same initial instrument, Andrews, Boyne and Enticott (2006) developed a different instrument to explain poor performance in English local authorities. The survey

instrument of authorities' characteristics totaled twelve items related to mismanagement. There were two items for 'performance management', one for 'clear priorities', two for 'internal partnership working', four for 'links between priorities and community needs', one for 'external partnership', one for 'managerial leadership', and one for 'political leadership'. Values for the internal validity of the indices were presented. The 'performance management' index had a Cronbach alpha of .96. One item had to be dropped from 'links between priorities and community needs' and principal components analysis had to be used by the authors to obtain a Cronbach alpha of .79 for 'links between priorities and community needs' and .69 for 'internal partnership working'. Naturally, for the other single-item indices, no further tests were needed.

The survey instrument of municipalities' internal characteristics has a total of fourteen items using four of the five categories described earlier. MAMROT and early survey takers asked that the 'external partnership' question, "this municipality welcomes partnership with the private sector", be dropped. For endogeneity reasons, the items related to performance management are not used to explain the use of performance information. There are two items for 'clear priorities', one for 'internal partnership working', four for 'links between priorities and community needs', one for 'managerial leadership', and one for 'political leadership' in this research.

On every item, surveyed takers could identify if the statements described their municipality, by agreeing with the statement on a four-point Likert scale. A full disagreement with the perception that the management characteristic applied in their

municipality was code ‘1’; a full agreement was coded ‘4’. Therefore, the clear priorities characteristics span from two to eight points, the links between priorities and community needs characteristics span from four to sixteen, and the internal partnership working characteristics, political leadership and managerial leadership span from four to sixteen.

Table 16. Descriptive Statistics of Independent Variables – Clear Priorities

Indicate your level of agreement regarding the following statements on the management indicators*, on the current situation in your municipality* *(1=disagree to 4=agree)*							
	Question		**n**	**min**	**max**	**mean**	**SD**
clear priorities	8		272	2	8	5.967	1.278
		Statements					
	8.1	There are clear links between the objectives and priorities of our service and those for the municipality as a whole	277	1	4	2.989	.810
	8.12	The municipality's objectives are clearly and widely communicated by managers of different services	282	1	4	2.979	.778

Table 17. Descriptive Statistics of Independent Variables - Links between Priorities and Community Needs

Indicate your level of agreement regarding the following statements on the management indicators, *on the current situation in your municipality*
(1=disagree to 4=agree)

Question			n	min	max	mean	SD
links between priorities and community needs	8		270	4	16	13.141	2.146
		Statements					
	8.3	The general manager and most managers place the needs of users first and foremost when planning and delivering services	285	1	4	3.540	.625
	8.4	Strategic planning is generally made in consultation with our external stakeholders	277	1	4	2.823	.835
	8.6	Working more closely with our citizens is a major part of our approach to service delivery	286	1	4	3.329	.663
	8.7	Citizens' demands are important in driving service improvement	283	1	4	3.481	.609

Table 18. Descriptive Statistics of Independent Variables - Internal Partnership Working – Political Leadership – Managerial Leadership

Indicate your level of agreement regarding the following statements on the management indicators, *on the current situation in your municipality* *(1=disagree to 4=agree)*							
	Question		**n**	**min**	**max**	**mean**	**SD**
internal partnership working	8		283	1	4	3.163	.759
		Statements					
	8.2	Co-ordination and joint working among the different municipal services is a major part of our approach to the organization of services					
political leadership	8		286	1	4	3.629	.558
	8.8	Political leadership is important in driving performance improvement					
managerial leadership	8		287	1	4	3.613	.580
	8.9	The general manager is important in guiding decision making to drive performance improvement					

The statistical internal reliability scores for Cronbach alphas for the two composite indices are respectively of .467, and .776. Yet again, a higher value of an index reflects a more intense perception of characteristics favorable to the use of performance measurement.

Independent variable: performance of the municipality

As it was explained earlier in chapter 2, in 2007 the number of indicators was reduced from nineteen indicators to fourteen indicators. These fourteen mandatory indicators form the basis on which the performance of municipalities can be assessed. Several precisions are needed when performance is being added to the model.

First, the fourteen mandatory indicators do not cover all municipal services. For once, public libraries, fire services and police services are not currently included in the list of mandatory indicators. Second, performance should be understood as relative performance. The portal where managers are invited to assess themselves consists of comparison of the municipality indicator value to the appropriate quartile values for municipalities of its own size. Other than the relative position for municipalities of that size, no judgment is offered by MAMROT in defining what performance is, and what differentiates good or bad, or even better from worst performance.

Beside a passing remark to the effect that performance usually translates as effectiveness and efficiency (MAMSL 2004, 3), nowhere in MAMROT's brochures and guides is better performance defined. In the online portal, where the comparative data are

presented to municipal managers, the fourth quartile represents the highest values for the indicators. For example, higher plowing cost and more frequent water boiling notices are in the fourth quartile. In order to be able to include performance into the model, the decision made in Charbonneau, Bellavance and Holzer (2010:13-15) was reproduced here. In the context of this research, it is hypothesized that cost should be lower rather than higher; incidence of negative events like boiling notices should be minimized; the debt should represent a lower percentages of assets rather than a higher percentage, etc. Thus, better performance can be defined as being in the first quartile (lower plowing cost and less frequent water boiling notices); worst performance can be defined as being in the fourth quartile (higher plowing cost and potential higher retirement rate). This evaluation by quartiles for municipalities of similar characteristics is precisely how Zafra-Gómez, López-Hernández and Hernández-Bastida (2009:157) evaluated the performance in one of their recent studies on Spanish municipalities. Relative performance becomes the proportion of times a municipality is featured in the first and fourth quartiles for the fourteen indicators for the 2007 and 2008 period.

Table 19. Descriptive Statistics – Higher Performance – Lower Performance

Independent variable		**n**	**min**	**max**	**mean**	**SD**
Higher performance		312	0	15	4.606	3.074
	Year					
	2008	312	0	9	2.240	1.661
	2007	312	0	8	2.365	1.654
Lower performance		312	0	15	4.026	2.956
	Year					
	2008	312	0	8	1.573	1.573
	2007	312	0	7	1.549	1.549

Independent variable: socio-demographic characteristics of the municipality

To follow the customs of previous research about performance measurement (de Lancer-Julnes and Holzer 2001, 695; Askim, Johnsen and Christophersen 2008, 303), some socio-demographic variables are included in the models. Initially, the size of the population, the size of the budget, the area of the municipality, and the administrative region where the municipality belongs were expected to be part of the model. However, results from collinearity diagnostics revealed that the variables 'size of population' and 'size of budget' are highly correlated. Municipalities with larger population tend to have larger budgets. One of the two variables had to be dropped. The 'size of population' variable has no missing value; the 'size of budget' has many. The values for 'size of population' variable were available for 2008; those for the 'size of budget' dated from 2006. The 'size of population' variable was kept while the 'size of budget' variable was dropped. Additionally, a Wald test was performed on the independent variables to determine if the coefficients are jointly statistically different from zero. They are. As a result, all variables were kept. Therefore, in this research, the population size (from the categories of table 6) and area of municipalities (in km^2, times 1000) are included.

Table 20. Descriptive Statistics – Size of Population – Size of Municipality

Independent variable	**n**	**min**	**max**	**mean**	**SD**
Size of population	312	1	9	2.766	1.759
Size of municipality	312	890	7923160	173739	505169.1

Data Analysis Procedures

The data from the returned electronic surveys was compiled by a research assistant at the Center for the Promotion of Municipal Management Excellency. From this excel spreadsheet, a STATA database was obtained and later used to perform statistical descriptive and regression analyses.

Method #2: Content Analysis

The answers to the two open-ended questions were codified by themes and ideas, not by sentences or comment. Thus, the total frequency of themes is greater than the number of comments. The results of this analysis are presented in chapter 4. Selected answers from participants of the focus groups are also presented and discussed in chapter 5.

Chapter 4: Quantitative Analyses

Analysis of the use of performance information

The descriptive statistics presented in chapter 3 reveal that the performance information carried through the management indicators is not widely used. Of the 312 respondents that are General Managers (out of the 390 survey respondents), 55.5% said that the performance indicators are never used in their municipalities. The rest of the General Managers say that performance information is used. Of these 39.5% are from municipalities where the information is seldom used; 4.5% are from municipalities where this information is used often, and only 0.3% (one manager) use it very often. The of managers either are from municipalities where performance information is perceived as not being used or only seldom used. The substantive nuance is thin. This can elevate the difficulty of finding differences between users and non-users of performance measures, as municipalities with no or low intensity of use can be similar. This constitutes the nuance problem. There is also a novelty problem. As mentioned earlier, the nuance problem is compounded by the fact that almost no study on the use of performance information utilizes regression analyses. There was some guidance in finding instruments to measure the factors influencing use. However, these independent variables were not fine-tune survey items for regression purposes. Most previous studies were content to report descriptive statistics, or offer simple correlation tables (Taylor, 2009:859-860). Nevertheless, logistical regressions performed in this study reveal that there are significant differences between municipal non-users and users of performance information.

In previous research, use of performance was theorized in different ways. Akin to first-generation studies on the use of performance, the general use of performance information is included in this study. Also, three specific uses of information were adapted from Rogers (2006): management use, budgeting use, and reporting use. Lastly, the actual use of performance information is also included. Following the descriptions from the literature presented in chapter 1, actual use refers to functions with demonstrable impacts on operations. Actual use comprises management and budgeting uses, and excludes reporting uses.

In total, there are five statistical analyses, one for every type of use. All analyses are logistic regressions with a binary dependent variable. Either there is no occurrence of a use, or there is at least some manifestation of it. For each analysis, six iterations are presented. The first four iterations, models 1 to 4, are partial models. These models are presented to reveal the robustness of the most important model: the fifth model. Model 5 is the full model that ought to have our attention.

As presented in chapter 3, there are instances of independent variables with missing data. The proportion of missing data can be in part explained by the fact that the independent variables are indices regrouping variables. By default, in the presence of missing data, STATA deletes cases in a list-wise fashion. As soon as one variable has incomplete information, the observation with that piece of missing data is deleted. This explains why model 5, the full model, has a smaller number of observations (227) than models 1 or 4, with a full 312 observations. A smaller number of observations means less statistical

power for the model. In the full model, it becomes more difficult to identify if differences are statistically significant. To remedy the situation, a sixth model is added to all five analyses. This sixth model is labeled “multiple imputations (5)” model.

Multiple imputations are based on the early work of Rubin (1987) and Li, Raghunathan and Rubin (1991). Statistically, “as with any finite population survey where valid frequency inference is desired from predictive procedures: (1) variables involved in the definition of estimands should be predicted, and (2) variables involved in the survey design should be used as predictors” (Rubin, 1996:478). The ‘multiple imputations (5)’ model presents the average values of five iterations of plausibly imputed data in lieu of missing values. The ‘multiple imputations (5)’ is a reiteration of the full model, only with the statistical power of 312 observations instead of 227. The ‘multiple imputations (5)’ model uncovers links that model 5 would have missed.

The results of the five analyses are presented in tables Z1 to Z5. Contrary to coefficients in OLS regressions, the raw coefficients of logistic regression cannot readily be interpreted. The coefficients of the independent variables are marginal coefficients. The multiple imputation option is a new feature offered in STATA 11. Currently, it is not compatible with many of the more advanced options for logistic regression. The calculation of marginal coefficients and the computation of model fit are not offered with multiple imputed logistic models. This is why they are not presented for the ‘multiple imputations (5)’ in tables Z1 to Z5. To ease the interpretation of the results presented in tables Z1 to Z5, the odds ratios are reported in the form of percentage.

Table 21. Logistic Regression Analysis of General Use of Performance Information (robust std. err.; marginal coefficients; odd-ratio in %)

	Model 1			Model 2			Model 3			Model 4			Model 5			Mult. Input.(5)	
	m. coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	OR %	sig.
1. Relative performance																	
Higher performance	0.032	***	14.1	0.023	**	9.7	0.034	***	14.5	0.028	***	12.1	0.015		*n.s.*	12.6	***
Lower performance	0.021	**	9.1	0.016		*n.s.*	0.013		*n.s.*	0.018	*	7.5	0.004		*n.s.*	*n.s.*	
2. Barriers																	
Unwillingness barrier	---			-0.079	***	-27.2	---			---			-0.082	***	-28.1	-26.8	***
Inability barrier	---			-0.015		*n.s.*	---			---			-0.003		*n.s.*	*n.s.*	
Prevented barrier	---			-0.022		*n.s.*	---			---			-0.034		*n.s.*	*n.s.*	
3.Internal characteristics																	
Clear priorities	---			---			0.025		*n.s.*	---			-0.007		*n.s.*	*n.s.*	
Links between priorities and community needs	---			---			0.040		*n.s.*	---			0.080		*n.s.*	*n.s.*	
Internal partnership working	---			---			-0.016		*n.s.*	---			-0.003		*n.s.*	*n.s.*	
Political leadership	---			---			-0.156	*	-46.6	---			-0.117		*n.s.*	*n.s.*	
Managerial leadership	---			---			0.156		*n.s.*	---			0.153	*	85.3	*n.s.*	
4. Socio-demo. characteristics																	
Size of population	---			---			---			0.024		*n.s.*	0.022		*n.s.*	*n.s.*	
Size of municipality	---			---			---			-1.97E-08		*n.s.*	-6.59E-08		*n.s.*	*n.s.*	
Pseudo R-squared	0.039			0.176			0.056			0.043			0.193			---	
Adj. count R-squared	0.115			0.325			0.150			0.086			0.385			---	
	n=312			n=253			n=257			n=312			n=227			n=312	

* p < .10 ** p < .05 *** p < .01

Table 22. Logistic Regression Analysis of Management Use of Performance Information (robust std. err.; marginal coefficients; odd-ratio in %)

	Model 1			Model 2			Model 3			Model 4			Model 5			Mult. Imput.(5)	
	m. coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	OR %	sig.
1. Relative performance																	
Higher performance	0.026	***	19.3	0.020	**	14.9	0.031	***	22.6	0.018	**	13.0	0.014		*n.s.*	13.7	**
Lower performance	0.019	***	13.7	0.010		*n.s.*	0.015	*	10.4	0.013	*	9.5	0.002		*n.s.*	*n.s.*	
2. Barriers																	
Unwillingness barrier	---			-0.042	***	-25.5	---			---			-0.046	***	-30.1	-27.5	***
Inability barrier	---			-0.005		*n.s.*	---			---			0.000		*n.s.*	*n.s.*	
Prevented barrier	---			-0.017		*n.s.*	---			---			-0.014		*n.s.*	*n.s.*	
3.Internal characteristics																	
Clear priorities	---			---			-0.002		*n.s.*	---			-0.017		*n.s.*	*n.s.*	
Links between priorities and community needs	---			---			0.003		*n.s.*	---			0.019		*n.s.*	*n.s.*	
Internal partnership working	---			---			-0.009		*n.s.*	---			0.003		*n.s.*	*n.s.*	
Political leadership	---			---			-0.109	**	-51.4	---			-0.074		*n.s.*	*n.s.*	
Managerial leadership	---			---			0.107	*	103.2	---			0.083		*n.s.*	*136.2*	**
4. Socio-demo. characteristics																	
Size of population	---			---			---			0.033	***	*25.2*	0.027		*n.s.*	*23.8*	***
Size of municipality	---			---			---			-4.05E-09		*n.s.*	-9.74E-09		*n.s.*	*n.s.*	
Pseudo R-squared	0.063			0.195			0.087			0.086			0.243			---	
Adj. count R-squared	0.016			0.161			0.037			0.127			0.204			---	
	n=312			n=253			n=257			n=312			n=227			n=312	

* p < .10 ** p < .05 *** p < .01

Table 23. Logistic Regression Analysis of Budgeting Use of Performance Information (robust std. err.; marginal coefficients; odd-ratio in %)

	Model 1			Model 2			Model 3			Model 4			Model 5			Mult. Imput.(5)	
	m. coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	OR %	sig.
1. Relative performance																	
Higher performance	0.032	***	15.9	0.019	*	8.7	0.034	***	16.5	0.031	***	15.1	0.013		*n.s.*	15.0	***
Lower performance	0.017	*	8.2	0.005		*n.s.*	0.011		*n.s.*	0.016		*n.s.*	0.002		*n.s.*	*n.s.*	
2. Barriers																	
Unwillingness barrier	---			-0.049	***	-19.5	---			---			-0.052	***	-21.2	-23.1	***
Inability barrier	---			-0.009		*n.s.*	---			---			-0.007		*n.s.*	*n.s.*	
Prevented barrier	---			-0.027		*n.s.*	---			---			-0.030		*n.s.*	*n.s.*	
3.Internal characteristics																	
Clear priorities	---			---			0.046		*n.s.*	---			0.031		*n.s.*	*n.s.*	
Links between priorities and community needs	---			---			-0.046		*n.s.*	---			-0.023		*n.s.*	*n.s.*	
Internal partnership working	---			---			-0.015		*n.s.*	---			-0.012		*n.s.*	*n.s.*	
Political leadership	---			---			-0.074		*n.s.*	---			0.050		*n.s.*	*n.s.*	
Managerial leadership	---			---			0.166	**	111.8	---			0.120		*n.s.*	132.0	**
4. Socio-demo. characteristics																	
Size of population	---			---			---			0.009		*n.s.*	0.011		*n.s.*	*n.s.*	
Size of municipality	---			---			---			-1.92E-08		*n.s.*	-1.98E-07		*n.s.*	*n.s.*	
Pseudo R-squared	0.042			0.124			0.063			0.043			0.154			---	
Adj. count R-squared	0.068			0.200			0.034			0.049			0.184			---	
	n=312			n=253			n=257			n=312			n=227			n=312	

* p < .10 ** p < .05 *** p < .01

Table 24. Logistic Regression Analysis of Reporting Use of Performance Information (robust std. err.; marginal coefficients; odd-ratio in %)

	Model 1			Model 2			Model 3			Model 4			Model 5			Mult. Imput.(5)	
	m. coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	OR %	sig.
1. Relative performance																	
Higher performance	0.022	**	10.2	0.016		*n.s.*	0.023	**	10.4	0.019	**	8.6	0.007		*n.s.*	*n.s.*	
Lower performance	0.011		*n.s.*	0.001		*n.s.*	0.009		*n.s.*	0.008		*n.s.*	-0.001		*n.s.*	*n.s.*	
2. Barriers																	
Unwillingness barrier	---			-0.062	***	-24.0	---			---			-0.063	***	-24.3	-22.5	***
Inability barrier	---			-0.008		*n.s.*	---			---			-0.003		*n.s.*	*n.s.*	
Prevented barrier	---			-0.018		*n.s.*	---			---			-0.030		*n.s.*	*n.s.*	
3.Internal characteristics																	
Clear priorities	---			---			0.001		*n.s.*	---			-0.031		*n.s.*	*n.s.*	
Links between priorities and community needs	---			---			0.053		*n.s.*	---			0.089		*n.s.*	*n.s.*	
Internal partnership working	---			---			0.027		*n.s.*	---			0.024		*n.s.*	*n.s.*	
Political leadership	---			---			-0.132	*	-43.0	---			-0.135	*	-45.2	*n.s.*	
Managerial leadership	---			---			0.130	*	73.9	---			0.132	*	80.0	*n.s.*	
4. Socio-demo. characteristics																	
Size of population	---			---			---			0.019		*n.s.*	0.017		*n.s.*	*n.s.*	
Size of municipality	---			---			---			-1.71E-08		*n.s.*	-1.67E-07		*n.s.*	*n.s.*	
Pseudo R-squared	0.009			0.134			0.04			0.022			0.156			---	
Adj. count R-squared	0.068			0.140			0.051			0.018			0.202			---	
	n=312			n=253			n=257			n=312			n=227			n=312	

* p < .10 ** p < .05 *** p < .01

Table 25. Logistic Regression Analysis of Actual Use (Management Use and Budgeting Use) of Performance Information (robust std. err.; marginal coefficients; odd-ratio in %)

	Model 1			Model 2			Model 3			Model 4			Model 5			Mult. Imput.(5)	
	m. coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	*m.* coef.	sig.	OR %	OR %	sig.
1. Relative performance																	
Higher performance	0.039	***	18.2	0.029	***	12.8	0.044	***	20.6	0.035	***	16.0	0.025	*	11.0	18.1	***
Lower performance	0.021	**	9.5	0.005		*n.s.*	0.014		*n.s.*	0.017	*	7.8	-0.002		*n.s.*	*n.s.*	
2. Barriers																	
Unwillingness barrier	---			-0.071	***	-25.8	---			---			-0.079	***	-28.4	-27.8	***
Inability barrier	---			-0.011		*n.s.*	---			---			-0.006		*n.s.*	*n.s.*	
Prevented barrier	---			-0.042		*n.s.*	---			---			-0.041	*	-15.9	-14.2	*
3.Internal characteristics																	
Clear priorities	---			---			0.035		*n.s.*	---			0.010		*n.s.*	*n.s.*	
Links between priorities and community needs	---			---			-0.066		*n.s.*	---			-0.047		*n.s.*	*n.s.*	
Internal partnership working	---			---			-0.003		*n.s.*	---			0.010		*n.s.*	*n.s.*	
Political leadership	---			---			-0.164	**	-50.2	---			-0.059		*n.s.*	*n.s.*	
Managerial leadership	---			---			0.216	***	150.4	---			0.199	**	131.7	179.5	***
4. Socio-demo. characteristics																	
Size of population	---			---			---			0.023		*n.s.*	0.027		*n.s.*	*n.s.*	
Size of municipality	---			---			---			-2.09E-08		*n.s.*	-1.02E-07		*n.s.*	*n.s.*	
Pseudo R-squared	0.054			0.188			0.084			0.059			0.227			---	
Adj. count R-squared	0.110			0.288			0.079			0.110			0.337			---	
	n=312			n=253			n=257			n=312			n=227			n=312	

* p < .10 ** p < .05 *** p < .01

The results from table 21 to 25 help us answer the first of our research questions by offering empirical ground in testing the first eight hypotheses. At this point, it is useful to reintroduce the first research question:

> **R1. Which factors account for the uses of performance measurement by municipal managers?**

Barriers to uses of performance information

The first three hypotheses are related to the perceived presence of barriers hindering the use of performance information. The expectations were that more barriers to the use of performance information would correlate with diminished uses of performance information.

> **H1. It is expected that managers who express their unwillingness to use performance indicators, will indeed use performance measurement less than managers that do not perceive this barrier.**
>
> **H2. It is expected that managers who express their inability to use performance indicators, will indeed use performance measurement less than managers that do not perceive this barrier.**
>
> **H3. It is expected that managers who express being prevented from using performance indicators, will indeed use performance measurement less than managers that do not perceive this barrier.**

In the survey, the barriers were listed in turns, in no specific order. This survey instrument about barriers was adopted with minimal adaptations from Siverbo and Johansson (2006:283-284). These authors regrouped the barriers in three groups: barriers related to the unwillingness to use performance indicators, barriers related to the inability to use them, and barriers related to being prevented from using them. In both Siverbo and

Johansson's (2006) study and the present one, survey takers were not cognizant of the thematic grouping of barriers.

In relation to the barriers hindering the general use of performance information, the results of table 21 for the full model and the imputed model are that only one of the barriers has a verifiable impact. The perceived barriers of being unable or prevented to use performance information do not seem to impact the general use of the management indicators. On the other hand, the barrier reflecting an unwillingness to use the indicators is statistically significant. With 99% confidence, we observe that when the values of all factors are average, an additional point on the twelve-increment scale of unwillingness will decrease the probability that a General Manager would express using the indicators by 26.8% to 28.1%.

The same pattern is found for the specific uses of performance information: the only statistically significant barrier is the one related to the unwillingness to use the indicators. The marginal effects of the unwillingness barrier on the management, reporting, and budgeting uses of performance information are comparable. With 99% confidence, we can say that when all the values of the independent variable are average, an additional point on the unwillingness scale decreases the probability that the General Manager is using the indicators for management, budgeting and reporting. The decreases are respectively of 27.5% to 30.1%, 21.2% to 23.1%, and 22.5% to 24.3%. At present, the average value of the unwillingness barrier score for the 227 municipalities of model 5 is 12.03. This means that the typical General Manager "somewhat agree" that the four unwillingness barriers are on average present in his/her municipality. The proportions of

those General Managers who report some management, budgeting and reporting uses are respectively of 15.2%, 32.3%, and 34.3%. For General Managers with an unwillingness score of 8, who on average "somewhat disagree" that the four unwillingness barriers are present, the proportion of uses respectively jumps to 41.1%, 55.5%, and 61.5%. The effect of this variable is sizable.

The perceived inability to use performance does not correlate with the actual use for performance. For the actual use of performance information, we observe with 99% confidence that the marginal effect of the unwillingness barrier decreases the probability of using indicators by approximately 28% (27.82% to 28.4%). For a General Manager, the effect of "somewhat disagreeing" instead of "somewhat agreeing" to presence of the unwillingness barriers translates into an occurrence of actual use of 70.6% instead of 38.5%. With less certainty (90% confidence), the perception of being prevented from using the indicators is linked to the probability of actually using the indicators. This is the only type of use where this influence is noticeably felt. A marginally stronger feeling of being prevented from using indicators decreases the probability that General Managers will actually use indicators by just about 15% (14.18% to 15.9%).

All in all, the results of the five analyses on the effects of the barriers reveal that the perceived inability to use the indicators does not impact the uses of performance information. We fail to reject the null hypothesis for H2.

Only one of the five analyses offers evidence to support the hypothesis that the perception of being prevented from using the indicators impacts their uses. The perception of being prevented from using indicators does not seem to have an impact on the general, management, budgeting, and reporting uses of performance information. Only when the definition of use is relaxed to include some expression of either management or budgeting uses (actual use), does the perception of being prevented manifest itself. When a correlation is found, it is significant at the 90% confidence level. The results are not statistically robust. Substantively, the influence is important; although not as important as the unwillingness barrier. The evidence to support hypothesis H3 is modest. With caution and reserve, we reject the null hypothesis for H3.

The case to point out the unwillingness barrier as a prime influence of performance information uses is strong. The manifestation of management behaviors stemming from an unwillingness to use the indicators is constant across the board. Once the discrete changes of variations of unwillingness are taken into account, it becomes clear that this barrier is key to understand why certain General Managers choose not to use the indicators in their decision making and operations.

Internal characteristics of the municipality

Variables accounting for internal characteristics of the municipality were included in the model. The five dimensions of municipalities' characteristics are imported whole from a British survey instrument utilized for government and academic research alike. The items for the five dimensions come from a larger list of that same instrument. Similarly to the

way barriers were measured, the score for the internal characteristics of municipalities reflects the perceptions expressed by General Managers in regard of their municipality. Again, survey takers were not aware of how the items would be grouped.

Here are the hypotheses H4 to H8:

> **H4. It is expected that managers who describe their municipality as having clear priorities, will use performance measurement more than managers that do not perceive their municipality this way.**
>
> **H5. It is expected that managers who describe their municipality as having internal working partnership, will use performance measurement more than managers that do not perceive their municipality this way.**
>
> **H6. It is expected that managers who describe their municipality as having links between priorities and community needs, will use performance measurement more than managers that do not perceive their municipality this way.**
>
> **H7. It is expected that managers who describe their municipality as being under strong political leadership in regards to performance measurement, will use performance measurement more than managers that do not perceive their municipality this way.**
>
> **H8. It is expected that managers who describe their municipality as being under strong managerial leadership in regards to performance measurement, will use performance measurement more than managers that do not perceive their municipality this way.**

Table 21 features the results for the general use of performance information. In relations to internal characteristics, the results of model 5 are that only one characteristic correlates with the general use of performance measurement. The relationship is significant at the 90% confidence level. This result is not corroborated by the imputed model. Only six of the 287 General Managers of model 5 do not perceive that managerial leadership is influential for performance information use. The average score for this characteristic is

3.61, mid-way between “agreeing” and “somehow agreeing” with the applicability of the statement in their municipality. The discrete effect of managerial leadership on general use is more easily understood in terms of comparisons. The probabilities of being a user of indicators for two typical and (statistically) undistinguishable General Managers would be of 36.8% for a General Manager who “somewhat agrees” on the role of managerial leadership vis-à-vis 51.1% for a General Manager who “agrees.”

Similar results come from tables 22 to 24 on the specific uses of indicators. According to the full models, perceptions of clear priorities, links between priorities and community needs, internal partnership working, and political leadership are not correlated with management and budgeting uses. The relationships between managerial leadership, and management and budgeting uses are not significant in the full model, but are significant in the imputed models. In both cases, the statistical power of the full model is not strong enough to uncover the relationship. In the imputed models for management use, we can affirm with 99% confidence that the marginal effect of managerial leadership decreases the odds of using indicators by 136.2%. The same can be said at the 95% confidence level for budgeting use; the odds of the marginal effect of the variable are 132.0%.

The correlation between internal characteristics of municipalities and reporting use is less conclusive than the management and budgeting uses. The influence of managerial leadership is only felt in the full model, and only at the 90% confidence level. Table 24 also bolsters the only significant occurrence of political leadership’s influence on any use. The relationship also proves to be present at the 90% confidence level. Moreover,

the polarity of the marginal coefficient is contrary to expectations: stronger perceived political leadership on performance management would translate into a lower probability of reporting results. Results from further modeling reveal that interacting strong political leadership with stronger performance or weaker performance than average (performance one standard-deviation away) barely impacts the probability of reporting uses. The facts that political leadership is, (a) only significant once in tables 21 to 25, and (b) is of reverse polarity to what expected, makes the interpretation of the result perilous. If this finding was to be exact, it would means that increased involvement of elected officials in performance management would reduce the probability to have the results of the indicators reported to stakeholders.

The findings of internal characteristics for the general use and the specific uses reverberate with the actual use of performance information. Increased managerial leadership influences positively the use of performance information, while the influence of other internal characteristics cannot be demonstrated.

Overall, the findings on the internal characteristics do not support the hypotheses H4, H5 and H6. We fail to reject the null hypotheses about the influence of clear priorities, links between priorities and community needs, and internal partnership working. Stronger political leadership, as perceived by the General Manager, would only affect the use of indicators for reporting. More intense leadership from elected officials about performance management would not impact most uses of indicators. Nonetheless, it could impact

reporting negatively. It is with caution and restraint that we reject the null hypothesis for H7, but only in the case of reporting use.

A similar argument to the one related to the unwillingness barrier can be offered for the managerial leadership. However, the argument has to be scaled down. This variable is present in enough of the uses to reject the null hypothesis for H8.

Other variables

Four variables that were included in the regression were not covered by hypotheses. These variables were related to socio-demographic characteristics of municipalities and the performance of these municipalities, as measured by the Municipal Management Indicators. A brief description of these variables is warranted.

Higher performance, conceptualized as the number of indicators for which a municipality is featured in the "top" quartile of other municipalities of comparable size, proved to be statistically significant for four of the five uses. The influence of higher performance was established for the general use, management use, budgeting use, and actual use of performance information. The size of the marginal effect is relatively small compared to that of the unwillingness barrier and managerial leadership. After controlling for other factors, a small but discernable pattern remains for performance: General Managers from municipalities where comparative performance is high tend to use the indicators more than others. The value of indicators as a source of information, is used more often in

municipalities where it can be seen as a vindication of encouraging results than when it suggests suboptimal results.

The size of a municipality, as measured in population and area, is not a statistically significant variable. When other characteristics of municipalities are taken into account, the rural character (proxy of area) of a municipality or the fact that it is a village instead of a city, fail to influence on the uses of the indicators.

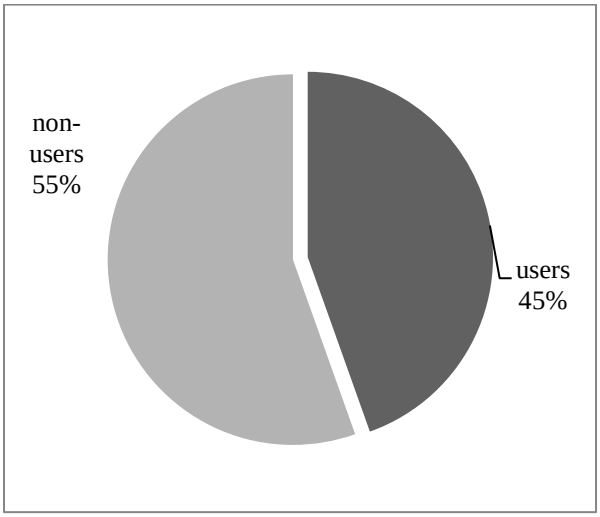
Now that the first research question has been answered, our analysis moves to the two remaining research questions tangential to the use of performance information by General Managers.

The second research question raised earlier in this study has to do with the comparison levels used by General Managers.

> **R2. What are the comparison levels being used by municipal managers in interpreting performance measures?**
>
> **H9. It is expected that, despite the availability of external comparative data, the most frequent comparison level used by managers will be historical comparisons on their own municipality.**

The relative frequencies presented in figure 5 below do not include all General Managers who responded to the survey. Rather, it presents the comparisons used by the 139 managers who do report using the indicators.

Figure 5. Relative Proportion of General Managers Reporting General Use of the Indicators



For example, the first box in table 26 should be understood as 72.66% of General Managers using the indicators compared the results of the ‘cost of the municipal roadway system, per km’ at least once. By definition, General Managers who do not use the indicators at all do not compare the values of the indicators for their municipality. If all managers were included in the table, the relative frequency presented in the table would be of 32.37%, not 72.66%.

Table 26. Frequency of Comparison Levels being Used by General Managers

Function & Activity	Indicator	Have you ever compared the results?	If your answer is *yes*, which elements were used while comparing your results?			
		Yes	Previous results	Quartiles from CPEGM* annual reports	Results from comparable municipalities	Others
Roads Municipal roadway system	Cost of the municipal roadway system, per km	**72.7%**	**64.0%**	**2.9%**	**17.3%**	**1.4%**
Roads Snow removal	Cost of snow removal, per km	**74.8%**	**62.6%**	**2.9%**	**20.9%**	**0.7%**
Public Hygiene Water supply, treatment and distribution	Number of breaks, per 100 km of pipes	**41,0%**	**39.6%**	**2.2%**	**10.1%**	**1.4%**
	Cost of distribution, per km of pipes	**43.9%**	**40.3%**	**2.2%**	**10.8%**	**1.4%**
	Cost of supply and treatment of water per m^3	**47.5%**	**41.7%**	**3.6%**	**12.2%**	**2.2%**
	Cost of water distribution per m^3	**46.8%**	**41.0%**	**3.6%**	**12.2%**	**2.2%**
Public Hygiene Used water treatment and sewage systems	Cost of treatment of used water per m^3	**45.3%**	**41.0%**	**2.2%**	**11.5%**	**0.7%**
	Cost of sewage system per km of pipes	**45.3%**	**41.0%**	**2.2%**	**10.1%**	**0%**
Global Financial Health	Percentage of debt service	**61.2%**	**51.8%**	**2.9%**	**25.9%**	**0.7%**
	Indebtedness percentage	**63.3%**	**52.5%**	**2.9%**	**27.3%**	**0%**
Human Resources	Training effort per employee	**36.0%**	**30.9%**	**1.1%**	**5.8%**	**0.7%**
	Percentage of training cost, compared to total payroll	**32.4%**	**28.8%**	**1.4%**	**5.8%**	**0.7%**
	Average length of health-related leaves of absence	**27.3%**	**24.5%**	**2.2%**	**4.3%**	**0.7%**
	Potential retirement rate	**27.3%**	**25.9%**	**1.4%**	**4.3%**	**0.7%**

CPEGM : *Centre de promotion de l'excellence en gestion municipale* [Centre for the Promotion of Excellence in Municipal Management] (http://neumann.hec.ca/cpegm/)

Until two years ago, the comparative results for the values of indicators were presented in reports prepared by CPEGM. In the summer of 2009, an interactive system in a password-protected web-portal took the relay. The "quartiles from CPEGM* annual reports" and "results from comparable municipalities" most probably measure the same concept of external comparisons presented by quartile. However, "results from

comparable municipalities" might also be construed as ad hoc comparisons with other municipalities by other means than MAMROT-sanctioned tools. However, for the purposes of testing hypothesis H9, this fact does not represent an impediment to the analysis.

The results of table 26 reveal that a majority of General Managers who use the indicators would compare their results for roads and financial indicators. Only a minority of General Managers would do the same with the rest of the indicators. The comparisons drawn by General Managers are overwhelmingly internal comparisons. The only dimension where the ratio of external-to-internal comparisons is less than two-to-one is for the financial indicators. It is clear from the results that the preferred comparison level of General Managers is their own performance over time. Hypothesis H10 is thus supported

Initially, one of MAMROT's objectives included the setting of targets for municipalities (MAMSL, 2004; MAMR, 2008). Target setting was voluntary. Sometime between December 2009 and July 2009, this official objective was discontinued by MAMROT (MAMR, 2009).

R3. How are targets set by municipal managers?

H10. It is expected that, despite the availability of external comparative data, the most frequent comparison level used by managers to set targets will be historical comparisons of their own municipality.

Table 27. Frequency of Targets Set by General Managers

Function & Activity	Indicator	Have you ever established targets?	If your answer is *yes*, which elements were used while establishing your targets?			
		Yes	Previous results	Quartiles from CPEGM* annual reports	Results from comparable municipalities	Others
Roads Municipal roadway system	Cost of the municipal roadway system, per km	**4.3%**	**4.3%**	**0%**	**0%**	**0%**
Roads Snow removal	Cost of snow removal, per km	**5.0%**	**4.3%**	**0%**	**0,7%**	**0,7%**
Public Hygiene Water supply, treatment and distribution	Number of breaks, per 100 km of pipes	**1.4%**	**2.2%**	**0%**	**0.7%**	**0.7%**
	Cost of distribution, per km of pipes	**0.7%**	**1.4%**	**0%**	**0%**	**0%**
	Cost of supply and treatment of water per m^3	**1.4%**	**1.4%**	**0%**	**0%**	**0%**
	Cost of water distribution per m^3	**0.7%**	**0.7%**	**0%**	**0%**	**0%**
Public Hygiene Used water treatment and sewage systems	Cost of treatment of used water per m^3	**0.7%**	**1.4%**	**0%**	**0%**	**0%**
	Cost of sewage system per km of pipes	**0.7%**	**1.4%**	**0%**	**0%**	**0%**
Global Financial Health	Percentage of debt service	**5.8%**	**2.9%**	**0%**	**3,6%**	**0.7%**
	Indebtedness percentage	**7.9%**	**3.6%**	**0%**	**4.3%**	**0%**
Human Resources	Training effort per employee	**2.9%**	**1.4%**	**0%**	**2.9%**	**0%**
	Percentage of training cost, compared to total payroll	**1.4%**	**1.4%**	**0%**	**1.4%**	**0%**
	Average length of health-related leaves of absence	**0.7%**	**1.4%**	**0%**	**0%**	**0%**
	Potential retirement rate	**0%**	**0.7%**	**0%**	**0%**	**0%**

It is obvious from the results of table 27 that General Managers who do use performance information prefer not to use targets. When targets are used, the levels of the targets are set with data from internal comparisons as opposed to external comparisons. The pattern is different for three of the indicators, two of them being the financial indicators. The differences in percentages in table 27 are very slight. They represent only one or two

General Managers. Literally, the most frequent comparison level used by managers to set targets is the historical comparison of their own municipality. However, given the slight differences in the incidences of comparison levels, it is dubious to suggest that these differences stand as strong evidence. The occurrence of targets is so rare and the differences are so slim that a definitive answer is impossible to reach. We cannot reject the null hypothesis for H10.

Now that the use of performance information by General Managers in Quebec has been charted, we can begin the analysis of sensemaking.

Analysis of sensemaking

The present study is interested in the perceptions of managers on how verdicts are reached with performance information. As we have seen earlier, sensemaking is an analytic activity that rests with comparisons of achieved performance with referent points. Sensemaking is an emerging topic in the performance measurement literature: few guidelines are available for the development of a survey instrument. Thus, the series of questions on sensemaking in the survey were designed for the needs of this study. This series of questions was asked to all survey takers, independently of their answer to question 1 concerning their overall use of performance indicators.

Use of performance measurement is a necessary but insufficient condition for sensemaking. Analysis of performance information is one step further than use, and at least two from symbolic use. In order to make the question applicable even to managers

who say they do not use performance information at all, the question was presented in the form of a scenario. Scenarios to assess sensemaking behavioral patterns in the face of failure and success were used once before, to study managers in the private sector (Wagner and Gooding, 1997: 280). The preamble to question 6 is:

> The following table presents a series of scenarios related to hypothetical results on a management indicator. Indicator values have to be interpreted in their context by managers: indicators do not speak by themselves. For each of the following scenarios described below, indicate how you would interpret the value of that indicator. There are no right or wrong answers. We are seeking your perceptions.

After the preamble, the survey question table 28 presents the survey instrument

Table 28. Dependent Variables related to Sensemanking of Performance Information

Dependent Variables	Survey Question	Citation
Sensemaking	Question 6 While analyzing the results of management indicators in your municipality, how would you interpret the following statements?	Developed for this study

The survey items come from referent points presented earlier in the first chapter. Referent points are listed for (1) last year's performance, (2) a predetermined target, (3) other municipalities of comparable size, and (4) other municipalities with comparable characteristics. A fifth item about professional standards was taken off the survey at MAMROT's request. There are three questions in relation to each referent point: a result *worse than* the referent point, a result that is the same/meeting/average the referent point, and a result *better than* a referent point. The terms *worse than* and *better than* were used, since the direction of indicators are different for cost indicators and effectiveness indicators. These questions are not specific to one indicator.

Some options are shaded in the tables. Initially, all four answers were available to all twelve items, independently of whether it made intuitive sense or not. For the first four items and the last four items, it is unlikely that a result worse than a referent point could be a success, and that a result better than a referent point should be seen as a failure. However, as we have seen in chapter 1, failure and success would not be symmetrical. The available answers for the four middle items are dissimilar; the answers for the four last items are more truncated than the first four items. The asymmetrical truncation of available answers does not invalidate the survey instrument, but it makes the analysis less straightforward than if all answers were available to all items.

The results of the sensemaking question are presented below. To facilitate comprehension, relative frequency is presented for all General Managers who did answer the question. Also, two relative frequencies are presented: one for managers who identified themselves as users of performance indicators at question 1, and the other for managers who did not. Perceptions of managers who do not use performance indicators are presented in brackets; perceptions of managers who do use performance indicators are on top in bold. The systematic difference of answers for each of the twelve items for those two groups, users and non-users, are presented in the last column.

Table 29. Relative Frequency of General Managers' Perceptions of Sensemaking, by Users and Non-Users of Performance Information

Scenarios	A success	Satisfactory	Un-satisfactory	A failure	*Chi²*
A result *worse than* last year performance?		**32.5%** [34.1%]	**65.1%** [59.5%]	**1.6%** [6.4%]	3.96 (Pr = n.s.)
A result *worse than* a predetermined target?		**25.0%** [24.6%]	**65.6%** [65.3%]	**9.4%** [10.2%]	0.05 (Pr = n.s.)
A result *worse than* other municipalities of comparable size.		**34.8%** [34.1%]	**64.3%** [59.9%]	**0.9%** [7.6%]	8.23 (Pr = 0.02)
A result *worse than* other municipalities with comparable characteristics.		**24.4%** [31.2%]	**71.3%** [57.6%]	**4.4%** [11.2%]	6.31 (Pr = 0.04)
The *same* result as last year performance?	**5.7%** [2.4%]	**88.5%** [82.4%]	**5.7%** [10.0%]	**0%** [4.8%]	9.48 (Pr = 0.02)
A result *meeting* a predetermined target?	**51.7%** [44.9%]	**48.3%** [55.1%]			1.09 (Pr = n.s.)
A result considered *average* to other municipalities of comparable size.	**4.3%** [8.3%]	**88.9%** [83.5%]	**6.8%** [8.3%]		1.86 (Pr = n.s.)
A result considered *average* to other municipalities with comparable characteristics.	**7.6%** [10.7%]	**86.6%** [81,2%]	**5.9%** [8.2%]		1.30 (Pr = n.s.)
A result *better than* last year performance?	**51.6%** [42.5%]	**48.4%** [57.5%]			2.03 (Pr = n.s.)
A result *better than* a predetermined target?	**69.0%** [60.0%]	**31.0%** [40.0%]			2.07 (Pr = n.s.)
A result *better than* other municipalities of comparable size.	**52.5%** [43.0%]	**47.5%** [57.0%]			2.19 (Pr = n.s.)
A result *better than* other municipalities with comparable characteristics.	**62.2%** [45.3%]	**37.8%** [54.7%]			6.77 (Pr = 0.01)

There are a few systematic differences between users and non-users of performance information. Overall, General Managers who do not use performance information do not express different perceptions from the ones who do.

After visual inspection of table 29, it becomes clear that answer patterns are different for the three groups of items: the *worse than*, the *same as*, and the *better than* groups. A classical hypothesis test, the one-sample mean-comparison test, can reveal whether General Managers favor satisficing over maximixing patterns in their sensemaking. By

coding failure and success as 1 for maximizing thinking, and by coding unsatisfying and satisfying as 0 for satisficing, one has to test if the mean is statistically different from 0.5. A sample mean statistically different from 0.5 would reveal that the General Managers significantly prefer one mode of thinking over another. The results of the t-tests are presented in table 30.

Table 30. Test of Mean on whether General Managers Think in Equal Proportions in Terms of Satisficing and Maximizing for Sensemaking, H_0: mean=0.5

Scenarios	*n*	mean	d.f.	student-t	|p-value|
A result *worse than* last year performance?	251	.040	250	-37.200	<.0001***
A result *worse than* a predetermined target?	246	.097	245	-21.229	<.0001***
A result *worse than* other municipalities of comparable size.	238	.050	237	-31.631	<.0001***
A result *worse than* other municipalities with comparable characteristics.	240	.079	239	-24.096	<.0001***
The *same* result as last year performance?	247	.065	246	-27.733	<.0001***
A result *meeting* a predetermined target?	236	.483	235	-0.520	0.607
A result considered *average* to other municipalities of comparable size.	238	.063	237	-27.682	<.0001***
A result considered *average* to other municipalities with comparable characteristics.	241	.091	240	-21.984	<.0001***
A result *better than* last year performance?	244	.471	243	-0.896	0.371
A result *better than* a predetermined target?	236	.644	235	4.613	<.0001***
A result *better than* other municipalities of comparable size.	241	.477	240	-0.708	0.480
A result *better than* other municipalities with comparable characteristics.	236	.538	235	1.173	0.242

*** p<.001,** p<.01, * p<.05

Going back to the fourth research question:

> **R4. While interpreting their own performance, when do municipal managers deem that the municipality's performance is (a) a failure, (b) unsatisfying, (c) satisfying, (d) a success?**

The results form tables 29 and 30 help us test hypothesis H12.

> **H11. It is expected that for sensemaking purposes, managers will reflect more in terms of satisficing than in terms of maximizing.**

Overall, the results from table 30 tell us that when it comes to sensemaking, General Managers think more in terms of satisficing than it terms of maximizing. Eight of the twelve items have statistical and substantial differences in the proportions of General Managers who consider results to be satisfactory or unsatisfactory, compared to results that are considered to be successes or failures. Hypothesis 11 is supported by the findings: General Managers do reflect more in terms of satisficing than in terms of maximizing.

The results are more telling if they are taken separately in terms of comparisons that are *worse than*, the *same as*, and *better than* the four referent points[ix]. For results that would be worse than the referent points, General Managers think significantly less in terms of failure than it terms of satificing. Intriguingly, nearly between a quarter and a third of managers even expressed the view that having results that would be *worse than* referent points could be satisfying.

Approximately the same could be said for comparisons that are the *same as* referent points: three of the four items have significantly more responses for satisfying than for maximizing. The only item where the difference is not large enough to be statistically significant is arguably of a different nature than the other item. Results that are either the

[ix] If the *worse than*, the *same as*, and *better than* items were indices, their Cronbach Alphas would respectively be of .81, .50 and .89.

same as last year's or of average value to other municipalities of comparable size or characteristics, can be thought of as being neutral. Meeting a target would have a more positive element to it: after all, by definition, targets are supposed to have a performance improvement element in them. Except for the items about meeting a target, where satisficing and maximizing are present in equal measures, the other neutral items are thought of more in terms of satisficing than in terms of maximizing.

As for items where results would be better than referent points, the general pattern observed before does not hold: General Managers arrive at maximizing verdicts as much as they do to satisficing verdicts. Again, the item about doing better than a target is different: General Managers are more likely to consider it a success than to consider it as just satisfying.

> **H12. It is expected that for sensemaking purposes, managers' verdict of performance will lean more in positive terms of satisfying and success than in negative terms of unsatisfying and failure.**

Very similar steps to the ones necessary to test H11 were done to test H12. With the difference that this time, positive answers like success and satisfying are coded 1, and negative answers like unsatisfying and failure are coded 0. Because the *better than* items do not have negative answers, they are not included in the series of t-test. Table 31 presents the results of the mean differences tests. Mean tests that are significant reveal that the proportion of answers that are either positive or negative are important enough that we can say it is different from neutrality (0.5).

Table 31. Test of Mean on whether General Managers Think in Equal Proportions in Positive and Negative Terms, H_0: mean=0.5

Scenarios	*n*	mean	d.f.	student-t	|p-value|
A result *worse than* last year performance?	251	.335	250	-5.540	<.0001***
A result *worse than* a predetermined target?	246	.248	245	-9.135	<.0001***
A result *worse than* other municipalities of comparable size.	238	.345	237	-5.036	<.0001***
A result *worse than* other municipalities with comparable characteristics.	240	.279	239	-7.611	<.0001***
The *same* result as last year performance?	247	.895	246	20.174	<.0001***
A result *meeting* a predetermined target?	*No negative answers for comparisons*				
A result considered *average* to other municipalities of comparable size.	238	.924	237	24.709	<.0001***
A result considered *average* to other municipalities with comparable characteristics.	241	.929	240	25.984	<.0001***
A result *better than* last year performance?	*No negative answers for comparisons*				
A result *better than* a predetermined target?	*No negative answers for comparisons*				
A result *better than* other municipalities of comparable size.	*No negative answers for comparisons*				
A result *better than* other municipalities with comparable characteristics.	*No negative answers for comparisons*				

*** p<.001,** p<.01, * p<.05

To test H12 on the positive and negative leaning of General Managers in sensemaking, the *same as* survey items are more revealing than the *worse than* items. After all, the *worse than* items should be perceived in negative terms. Results that would be either the same as last year's performance or average to other municipalities of comparable size or characteristics, are not seen in neutral ways. Neither are General Managers evenly split between those who see neutral results as positive and negative. On the contrary, the perceptions are very one-sided. In accordance with H13, sensemaking leans more

positively than negatively. This holds true for neutral results, which are more telling. Hypothesis H12 is supported by the data.

This tendency also manifests itself for negative (*worse than*) results. Unsurprisingly, the General Managers' sensemaking verdicts have a negative connotation for all *worse than* items. As previously mentioned, and somewhat surprisingly, between a quarter and a third of respondents consider that hypothetical results worse than the referent points would still be positive (satisfying).

> **H13. It is expected that for sensemaking purposes, managers will be more likely to interpret given performance levels in ways that put their performance in the best possible light.**

In this section, so far, we see that H11 and H12 are supported. General Managers do refer to performance more in terms of satisficing than in terms of maximizing. Additionally, when the proportion of General Managers using maximizing labels for judging results is indistinguishable from the proportion of General Managers using satisficing labels, it is in the favor of success, not failure. Also, General Managers' verdicts on results lean in positive rather than in negative terms. These three behavioral patterns, taken together, offer an argument in support to H13.

Further analysis can be added to make a stronger case for H13. By crossing over the answers of General Managers for symmetrical survey items, we can see if General Managers are consistent in their answers, or whether they indeed tend to interpret results in self-serving ways. For example, by verifying how one survey taker answered the "a result *worse* than other municipalities of comparable size" and the "a result *better* than

other municipalities of comparable size", which are polar opposites, we can assess the presence/absence of a self-serving bias. Tables x to x present a cross-table of results for the *worse than* and *better than* items for yearly variation of performance, target attainment, other municipalities of comparable size, and other municipalities of comparable characteristics.

Table 32. Cross-Tabulation of Municipal Managers' Sensemaking in Terms of Yearly Variation of Performance

Performance…		**…*better than* last year**				
		Failure	*Unsatisfactory*	*Satisfactory*	*Success*	Total
…*worse than* last year	*Failure*			1.7% (4)	0.4% (1)	2.1% (5)
	Unsatisfactory			27.5% (66)	35.8% (86)	63.3% (152)
	Satisfactory			22.9% (55)	11.7% (28)	34.6% (83)
	Success					
	Total			52.1% (125)	47.9% (115)	100.0% (240)

Pearson chi^2=12.88203 Pr=0.002

Table 33. Cross-Tabulation of Municipal Managers' Sensemaking in Terms of Target Attainment

Performance…		**… *better than* a predetermined target**				
		Failure	*Unsatisfactory*	*Satisfactory*	*Success*	Total
… worse than a predetermined target	*Failure*			1.7% (4)	6.4% (15)	8.1% (19)
	Unsatisfactory			18.8% (44)	47.4% (111)	66.2% (155)
	Satisfactory			15.0% (35)	10.7% (25)	25.7% (60)
	Success					
	Total			35.5% (83)	64.5% (151)	100.0% (234)

Pearson chi^2=18.8255 Pr=0.000

Table 34. Cross-Tabulation of Municipal Managers' Sensemaking in Terms of other Municipalities of Comparable Size

Performance…		…*better than* other municipalities of comparable size				
		Failure	*Unsatisfactory*	*Satisfactory*	*Success*	Total
… *worse than* other municipalities of comparable size	*Failure*			1.3% (3)	1.8% (4)	3.1% (7)
	Unsatisfactory			28.0% (62)	34.1% (78)	61.1% (140)
	Satisfactory			22.7% (52)	13.1% (30)	35.8% (82)
	Success					
	Total			51.1% (112)	48.9% (117)	100.0% (229)

Pearson chi^2=7.7684 Pr=0.021

Table 35. Cross-Tabulation of Municipal Managers' Sensemaking in Terms of other Municipalities with Comparable Characteristics

Performance…		…*better than* other municipalities with comparable characteristics				
		Failure	*Unsatisfactory*	*Satisfactory*	*Success*	Total
… *worse than* other municipalities with comparable characteristics	*Failure*			1.8% (4)	4.4% (10)	6.1% (14)
	Unsatisfactory			25.9% (59)	39.5% (90)	65.4% (149)
	Satisfactory			18.0% (41)	10.5% (24)	28.5% (65)
	Success					
	Total			45.6% (104)	54.4% (124)	100.0% (228)

Pearson chi^2=11.8037 Pr=0.003

To understand the last four tables, it helps to think of what it means to have two answers fitted in one of the six squares below. For example, the most frequent answer (39.47%) from General Managers is that a result *better than* other municipalities with comparable

characteristics is a success, whereas a result *worse than* other municipalities with comparable characteristics is unsatisfactory.

Table 36. Labels of Cross-Tabulation of Municipal Managers' Sensemaking

		…better than last year/a predetermined target/other municipalities of comparable size/other municipalities with comparable characteristics			
		Failure	Unsatisfactory	Satisfactory	Success
…worse than last year/a predetermined target/other municipalities of comparable size/other municipalities with comparable characteristics	Failure			**Pessimistic analyst** (consistent- mixed)	**Maximizer epitome** (consistent – maximize)
	Unsatisfactory			**Satisficer epitome** (consistent - satisfice)	**Optimistic analyst** (consistent – mixed)
	Satisfactory			**Biased analyst** (inconsistent – satisfice)	**Self-congratulating analyst** (inconsistent – mixed)
	Success				

In parentheses, the first element (before the hyphen), refers to the consistency that the *better than* item should receive a positive label and the *worse than* item should receive a negative label. The second element in parentheses (after the hyphen) is related to the maximizing/satisficing combination of answers. 'Mixed' refers to choosing one maximizing term with one satificing term; 'satisfice' means that both terms are satificing, and 'maximize' means that both terms are maximizing. In combination, it characterizes the perceptions of General Managers about their views of what it means for a result to be *better than* and *worse than* a given referent point.

The six types of analysts can be categorized in three groups: (a) one where verdicts for *worse than* results are harsher than *better than* results, (b) one where verdicts for *worse than* results correspond to the verdicts for *better than* results, and (c) one where verdicts

for *worse than* results are more lenient than *better than* results. The first group (a) contains one square, and represent only between 1.31% and 1.75% of General Managers. This group of pessimist analysts is very marginal. The second group (b) can be found in two squares that represent between 25.21% and 30.27% of General Managers. This group of satisficer and maximizer epitomes represents the minority of General Managers in Quebec. The third group (c) encompasses three squares and represents between 67.98% and 73.08% of General Managers. This group of optimistic, biased, and self-congratulating analysts represents the majority of General Managers in the Municipalities of Quebec.

All in all, the information cross-tabulated shows that the majority of managers do not judge results that are polar opposites in the same way. Between two-thirds and three-quarters of General Managers arrive at more lenient verdicts for negative results than for positive results. Furthermore, for between 25.64% and 35.81% of General Managers, *better than* results are positive, and *worse than* results cannot be negative. For a quarter to a third of General Managers, there is just no way that results from performance indicators could be unsatisfactory or failures. From the results presented earlier, General Managers think in negative satisfying terms for negative results, in positive satisfying terms for neutral results, and in positive satisfying to positive maximizing terms for positive results. It is safe to say that H13 is supported by the evidence. General Managers in Quebec interpret given performance levels in ways that put their performance in the best possible light.

Discussion of the quantitative data

Use of performance information

Do General Managers use performance information in Quebec? Do they use it them a lot? As we covered in chapter 1, the question is difficult to answer, because 'use' is measured in different ways in different studies. Findings from previous studies are that use, however defined, usually does not go further than 50%. Less than half of the departments within a city, less than half of decisions, less than half of management functions integrate information from performance measures. In Quebec, 44.55% of General Managers say that the indicators are used at least sometimes. In addition, 37.82% identify at least one management or budgeting function for which performance information would be enshrined. The proportions are somewhat comparable to large municipalities in national studies in the United States. The proportions are much lower than for Swedish municipalities in the voluntary RPE system. Siverbo and Johansson (2006:278) classified nine out of ten Swedish municipalities as a 'high-intensity' user. The same could be said for about one municipality out of twenty in Quebec.

The results of the quantitative analysis of uses are that there are two constant influences explaining the different uses of performance information. The first of these influences is the unwillingness of managers to use indicators. The barriers related to the perception of an inability to use data and being prevented to use data could not be identified as statistically significant forces. The second of these influences is the managerial leadership in terms of performance management. Managerial leadership in performance management demonstrably influences different uses of performance information.

However, when discrete changes on the probability of using indicators are factored in, the impact of managerial leadership is lesser than the impact of the unwillingness barrier. In discrete terms, the size of the independent influence of the unwillingness barriers outweighs the size of the independent influence of managerial leadership, by a factor of at least two/three to one. Political leadership would only be felt on the reporting use of performance information, where stronger involvement from elected officials would translate into more limited reporting. The influence of other internal characteristics, perceptions of clear priorities, links between priorities and community needs, and internal partnership working could not be demonstrated.

Interestingly, municipalities which tend to feature higher relative performance report higher uses of the information. It seems that municipalities with superior relative performance use performance information in their operations more than other municipalities, even after controlling for other factors. An alternative explanation could be that the use of performance measurement in a municipality in Quebec contributed to the achievement of better performance. At present, we cannot rule out this reverse causation. Lastly, despite frequent claims from managers that will be presented in chapter 5, it could not be demonstrated that the size of municipalities and the area of municipalities (a proxy for rural, suburban/urban municipalities) impacts the different uses of performance.

What we learned from the quantitative analyses of uses of performance information is that the factors correlated with use are linked with managers. Use is not due to outside

influences or immutable characteristics, such as being a very small municipality, or a rural municipality. A sizable portion of variation of general use can be explained by two factors: the unwillingness of managers to use the indicators and the leadership of managers. Taking other factors into account, the probability that a General Manager with a minimal unwillingness score and a maximal leadership score would be using the indicators is of 93.87%. The probability for his/her counterpart with a maximum unwillingness score and a minimal leadership score to use the indicators is of 4.37%.

In municipalities in Quebec, there are few General Managers who take the time to compare the results of the indicators to some referent points. When comparisons are made by General Managers, it is mainly to the past performance of their municipalities. This is the most basic level of possible comparison. Comparisons with referent points outside of the municipality are seldom used. One of the main benefits of a benchmarking system, the presence of comparative data from standardized indicators, is unutilized.

Targets are the vehicle to potential performance improvement. MAMROT never made the setting of targets mandatory. MAMROT recently abandoned the setting of targets as an official objective. General Managers are mandated to collect and transmit information on their activities, and most of them comply. General Managers are not mandated to set goals for themselves. It was suggested to them for most of the life of the Municipal Management Indicators. Only the rarest exceptions take the time to set goals for their municipality. Comparing results and setting targets are the most basic actions that precede an analysis of performance data. At present, in Quebec, it is difficult to

understand how General Managers are able to fully comprehend their performance when the values of indicators are to be thought of, in the absence of a goal.

Sensemaking

If we are to trust research on organizational behavior that was presented in chapter 1, discussions about corrective actions are likely to happen only once underperformance is formally recognized. An agreement that the observed performance is a failure is needed to instigate a diagnostic of operations, which possibly can result in corrective actions. A satisfying verdict on performance, or even a success, means that operational changes are unlikely to be pursued. In the public sector, at the municipal level, how do managers know that their municipality's performance is (a) a failure, (b) unsatisfying, (c) satisfying, (d) a success? What would constitute a success, a failure, satisfying or unsatisfying performance? To identify how performance information is calibrated by managers, questions were asked in the form of scenarios. They are scenarios in two ways. First, they are scenarios in the way the sensemaking series of questions was set up. The comparisons to referent points were introduced in scenarios that were qualified as being hypothetical. Second, in the way that most General Managers indicated earlier in the survey that they never compared the values of indicators. The series of questions was hypothetical indeed.

The results on sensemaking confirmed the hypotheses. All three hypotheses were supported by the data. General Managers would reflect more in satisficing terms than maximizing terms, unless performance would be better than a referent point. When this is

the case, then the proportion of a success verdict is roughly equal to a satisfying verdict. General Managers do lean more positively than negatively in their sensemaking. The majority of General Managers in Quebec, by a proportion from two-to-one to three-to-one, judge a result worse than a referent point in a more lenient way than a result better than that same referent point. This holds for comparisons with last year's performance, with targets, with other municipalities of comparable size, and with other municipalities of comparable characteristics.

In relation to these referent points, General Managers expressed in the survey that they do not think as *homo economicus*: rare are the General Managers who think in maximizing terms of success *and* failure. Across all four referent points, Herbert Simon's *satisficing man* is the second most common ideal type. The frequency pattern is the same for all four referent points. The most common inclination for General Managers in Quebec is to be an optimistic analyst, and think about *better than* and *worse than* respectively as a success and as being unsatisfying. The second most common behavior, as we just mentioned, is to consider performance solely in terms of satisfying and unsatisfying. Almost as frequently, the third shared proclivity is to show some biases in the analysis and consider *better than* and *worse than* performance the same way: as being satisfying. The next most common behavior is a stronger variant of the biased analyst: the self-congratulatory analyst. According to this kind of analyst, *worse than* results are satisfactory and *better than* results are successes. For this kind of General Managers, there are no occurrences of underperformance, no matter what the values of the indicators might suggest. Far behind, for three out of the four referent points, are managers who

think like *homo economicus*; and the least common is to be a pessimistic analyst who thinks in terms of failure and unsatisfactory results.

What, according to General Managers, would constitute a failure? Their response was: almost nothing. In the views of managers, 9.38% of self-proclaimed indicator users (and 10.17% of non-users) agree that a result worse than a predetermined target could be labeled as a failure. MAMROT does not provide sensemaking support of any sort to help managers in understanding what the performance means. The working assumption, after successive scaling back of official objectives, is that municipal managers can arrive at a verdict by themselves. This requires managers to perform analytical tasks. The results of the sensemaking show that left to themselves, General Managers are (or would be) partial in their analyses. The most blatant example is the sensemaking from comparisons with municipalities of similar size. In a directionless pure positional system, the chances of being "above[x]" or "below" average on a given indicator are strictly of 50%. Nevertheless, being "above[xi]" this level is seen in much more positive lights. The same seemingly incongruent thinking pattern applies to different levels for the other three referent point.

The survey data on sensemaking show that when managers must analyze performance, even in a hypothetical scenario bereft of consequences, they arrive at verdicts that are systematically optimistic, if not self-serving. In a satisficing mindset, neutral or average

[x] Again, MAMROT does not define what is "above" or "below" average. There are averages, and there are values on each side of them.

[xi] Again, MAMROT does not define what is "above" or "below" average. There are averages, and there are values on each side of them.

results are touted as positive. While few negative performance levels are seen as failures, thresholds for successes are much easier to reach.

In chapter 1, the themes of risk adversity and mistake avoidance were introduced. These themes would be the differences, some say the main differences, between managing in the public and private sectors. Managers in the private sector would be allowed to take risks, which means they are allowed to make mistakes, and would be judged on the bottom line. The bottom line takes the form of a budgetary surplus, supposedly in the presence of competition. However, in the public sector, there is no bottom line. By design, there is often no competition for a public agency. However, there is competition for elected officials. The party or parties in the opposition debate with elected officials about what a bottom line would look like. Elected officials will focus on positive results and the opposition will focus on negative results. As we have seen in chapter 1, it is largely accepted by both academics and practitioners that failure trumps success. Some even say that a failure negates successes. This means that elected officials and the managers working under them have incentive to minimize perceived failures rather than maximize perceived successes. Managers, and indirectly elected officials, would be judged on the presence of failure(s). More studies would be needed to prove this point, but mistake avoidance might explain why, for approximately a third of General Managers in Quebec, there is just no plausible scenario in which their performance could be labeled negatively.

Chapter 5: Qualitative Analyses

Analysis of the use of performance information

This chapter contains the qualitative analyses of the present research. This qualitative approach is used to support and complement the analyses of chapter 4. The data originates from two different sources. The first source of data comes from the two main open-ended questions of the survey. The second source of data comes from the focus groups organized by the Partners on Municipal Management Indicators Committee, which took place on January 28th, 2010.

Before moving forward in this section, a limitation has to be acknowledged: at present, only the author of this research coded the comments. Eventually, when the comments will be used for academic publications, at least one more coder will be needed to validate the author's work in coding the comments. Indeed, to be reliable, coding should show overlap and minimal variations between at least two coders. A measure called Kappa exists to assess inter-coder reliability. This measures the proportion of net agreement, once random agreement is excluded. The respondents' comments, just as the rest of the survey, were expressed in French. For the future, an additional coder with working knowledge in both performance measurement and the French language will have to be hired. Since only the author coded the data, for the immediate supportive use of qualitative methods in this research, the results should be considered preliminary and not definitive.

Comments from Managers explaining why they do not use the management indicators

The first question in the survey asked respondents if overall, the mandatory performance indicators: a) are used very often; b) are used often; c) are seldom used; d) are not used. The respondents were notified that if they answered "are not used", they were invited to briefly indicate why. Out of the 212 respondents who initially declared they do not use management indicators at all, 187 (88%) took the opportunity to explain why it is the case in the open-text comment section. A content analysis of their replies was done. The units of analysis for the content analysis were ideas. It became evident that many comments expressed multiple ideas in reply to that question. Every single comment was coded. No previous coding scheme was developed prior to the data collection.

The most frequent recurring theme on the reasons why managers do not use the management indictors is the perception that the mandatory management indicators are useless, or at least of no interest. This was by far the most recurrent theme. This can be appreciated in table 37, where the frequencies of themes are displayed.

Table 37: Reasons Given Why Management Indicators Are Not Used

Reasons (grouped by themes)	Frequency
Useless / do not see usefulness / no interest (managers)	66
Inherent difficulties for small municipalities	39
No interest from elected officials / incomprehension from mayor and-or council members	38
No time available	16
Lack of expertise / lack of personnel with expertise / lack of personnel	14
Simplistic / not realistic / not valid	14
No comparable municipality / impossible to compare municipalities with each other	13
Do not know what to do with performance information / difficulty in the interpretation of performance information (managers)	13
Limited use (present / past)	11
Maybe use indicators in the future	9
Accounting problems / cost accounting problems	8
Chaos	7
Difficulties based on indicators' churning	6
Difficulties with comparisons (software limitations)	5
Long delays in obtaining the data	2
Recognize some usefulness in indicators	2
Total (number of reasons from 187 respondents who say they do not use indicators) :	263

The comments about the perceived uselessness are rather straightforward and unambiguous. Some examples illustrate a shared feeling from managers in Quebec who do not use the management indicators.

> "These are not information that we use in our daily workday" [comment a23]
>
> "No one works with this information" [comment a132]
>
> "This information is not useful for us and I am under the impression that I waste precious time filling it out, let alone go back and consulting them. I have not yet seen the usefulness of this information we have to compile" [comment a26]
>
> "I compile them every year since we have the obligation to do so. I have to tell you that they [the indicators] only add to the considerable number of numbers on paper. They are not used" [comment a49]

> "This document on the management indicators has never been filled out by the city since 2003[xii] [comment a68]
>
> "We fill out the Ministry's annual survey [referring the mandatory transmission of data to MAMROT], but after that, the municipality does not refer to it. For us, it is not considered as a tool" [comment a121]
>
> "It does not mean anything" [comment a172]
>
> "I have been in this municipality for 1½ years and I never referred to them to justify a project" [comment a59]

Some participants expressed that management indicators are not useful, that budgeting tools are superior, and that performance indicators do not supply better information.

> "We already have other ways internally enabling us to evaluate our strengths and weaknesses; for example, prior years' spending. The indicators usually confirm what we already know" [comment a105]
>
> "We work instead with our financial reports to guide us and make necessary further corrections according to our situations along the years" [comment a108]
>
> "We compare ourselves from year to year with our numbers in the financial statements and we take note of variations in the budget – no need for another tool. (…)" [comment a131]

The comments above illustrate that some managers are indifferent and sometimes hostile to newer management tool like performance measurement. Some of these managers that do not use performance indicator openly state that they prefer to solely use other management tools like budgets.

The second most common theme among managers that do not use performance information is that there are inherent difficulties for small municipalities to use the

[xii] After verifications, there do not seem to be performance information for that municipality in MAMROT's database for 2006, 2007 and 2008.

management indicators in their operations. Some of the comments simply state the perception that performance indicators are ill suited for small municipalities.

> "It is very simple, the municipality is very small and he [not specified] does not feel the need to compare to other municipalities. The budget is done from year to year according to the current needs." [comment a18]

> "Complex and difficult to apply in a rural context (small municipality)" [comment a28]

> "There are only three employees and it is always a race, the budget is really carried through and we do [what we can] with our workforce" [comment a36]

> "The municipality is so small that I only work 2 days/week, thus I never used this document at present" [comment a65]

> "Possibly by lack of knowledge but also because municipal administration is very simple in a small village; as always, we find our answers elsewhere." [comment a153]

Some mentioned that performance indicators are not needed in small municipalities because they think important performance improvements cannot happen in such a context.

> "The indic. are not always realistic in small mun. The former do not have the budgets [resources] of cities, thus it is difficult to achieve large improvements. However, comparisons help to explain certain decisions." [comment a72]

This theme about the incongruence between performance indicators and small municipalities is seldom found in the literature. As we mentioned earlier, most North American studies are done with a sample of medium-sized municipalities (25K+ population) and little is known on how small and very small municipalities use, or in the present case cope, with performance measurement.

The other widespread theme emerging from the comments of managers who do not use indicators is that there is little or no interest from elected officials. Municipal elected officials in Quebec include the Mayor and council members. Here are some selected comments to that effect.

> "Notwithstanding explanations [from managers], the Council do not use these data" [comment a84]
>
> "Even though they [the indicators] are compiled every year, no council members consult them" [comment a95]
>
> "No relationship with the municipality's management and financial planning which is largely a political choice prioritize by elected officials" [comment a15]
>
> "Elected officials have been here for many years and trust their experience to make decisions" [comment a63]
>
> "Because elected officials cannot figure it [the indicators] out. They make the budget with the revenues, the spending, and the taxing rate are enough for them!" [comment a136]
>
> "Every year I transmit a copy of the management indicators to elected officials, and they [the indicators] never raised interest (not a question) from them [elected officials]" [comment a185]

There are examples where the impediment of small size is compounded with uncooperative elected officials.

> "Because we are a small municipality and Mister Mayor does not use this method because it is too complicated" [comment a140]

There are also examples where both managers and elected officials are allegedly opposed to the management indicators. The most explicit comment is:

> "Useless uselessness… elected officials do not know that it [the indicators] exists and do not want to know about it, and for public servants, this task is useless and does absolutely not serve any purpose except clutter the file cabinet" [comment1a73]

Out of the most frequent themes in the comments from managers who do not use the management indicators, are that the perceived uselessness from managers and for politicians. These themes are not new. Among other thing, they are included in the list of barriers included in question 7 of the survey, under ‘management indicators are not considered useful’ and ‘our elected officials are uninterested in the management indicators’. The list of barriers comes from another survey of municipalities in Sweden (Siverbo and Johansson, 2006).

An interesting theme that got some traction with managers who do not use performance information is the perception that it is impossible to compare municipalities in terms of service delivery. Comments to that effect took the following forms:

> “The indicators were invited by theoreticians (…) and it is impossible to compare 2 municipalities together” [comment a76]
>
> “(…) moreover, every municipality is unique. We cannot always make adequate comparisons” [comment a77]
>
> “Do not understand the usefulness of comparisons with our neighbors” [comment 98]
>
> “(…) Comparisons are always mistaken” [comment a118]
>
> “Not concrete enough: too many differences between municipalities” [comment a142]
>
> “For internal use only, because two municipalities cannot be compared (…)” [comment a159]

This idea that municipal benchmarking as a concept is misguided is unexpected and will be discussed in the ‘analysis of the qualitative data’ section. However, the most unexpected aspect of the qualitative part of the survey is found in the second main open-ended question.

Comments from managers regarding management indicators

At the very end of the survey, all respondents were given an opportunity to offer general comments regarding the management indicators: “If you have comments to formulate regarding the management indicators or the present survey, please add them in the space below.” A total of 152 respondents, or almost 40%, seized the opportunity and offered comments. Again, no prior coding scheme was prepared prior to the analysis of the data. There were no definite expectations on what respondents would say, as they could comments on any aspects of the management indicators. The nature of the comments from this general open-ended available to all managers is identical to the comments from the question on why some managers do not use the indicators. Only four additional categories from the initial coding scheme were needed to code the comments from the second open-ended question. Moreover, these four categories captured a limited number of comments. All in all, the general comments were in fact reiteration of why the indicators are useless, to complicate to use, simplistic and impractical. The main difference this time was that managers who say they use the indicators also expressed themselves. Table 38 summarizes the recurring theme in the comments for that question.

Table 38: General Comments Formulated about the Management Indicators by Users and Non-Users of the Management Indicators

Comments	Frequency	
	Non-Users	**Users**
Inherent difficulties for small municipalities	37	13
Useless/ do not see usefulness / no interest (managers)	21	8
No time available	16	9
Simplistic / not realistic / not valid	10	8
No interest from elected officials / incomprehension from mayor and-or council members	15	2
Some questions do not apply to our situation / perceived problems with the survey	12	5
No comparable municipality / impossible to compare municipalities with each other	8	6
Do not know what to do with performance information / difficulty in the interpretation of performance information (managers)	6	3
Need for training / request for training	7	2
Do not know what to do with performance information / difficulty in the interpretation of performance information (managers)	4	4
Maybe use indicators in the future	3	5
[positive comment]	2	13
Difficulties based on indicators' churning	1	5
Difficulties with comparisons (software limitations)	1	3
Accounting problems / cost accounting problems	1	3
Indicators do not take into account services provided by contractors	3	2
Long delays in obtaining the data	1	0
Chaos	0	1
We have our own indicators	1	0
Miscellaneous / [not really related with the indicators]	5	3
Total (number of reasons from 99 respondents who say they do not use indicators and 53 respondents who say they do use them) :	154	95

Table 38 indicates that self-declared users and non-users who took the time to comment on the indicators perceive multiple problems and impediment to their use. Out of the 249 themes that were present in the 152 comments, only fifteen comments were positive about the indicators. Two of the three widespread themes showed earlier are repeated. The lack of time seems to come up more often in this question than the lack of interest from elected officials. There are differences in the repartition of themes on the obstacles

and impracticalities of indicators from users and non-users. However, as we mentioned, there is a consensus about common perceived problems.

The exact same analysis from the previous section could be repeated here. One of the differences would be that the comments for this section would be lengthier. Once again, the idea is present that there is no time/room/usefulness from the management indicators because of the budget. Instead of presenting frequent comments that would reverberate with the comments that were presented above, selected comments that are substantially interesting will be presented.

One of the comments that is worth being reproduced was formulated by an individual with more than 20 years of experience as a dual General Manager–Chief Financial Officers in a municipality of less than 2,000 residents.

> "The lack of time and personnel create difficulties for me to fill the survey [referring the mandatory transmission of data to MAMROT] within the time delays. When we [unspecified] want to represent the real picture of what is going on in a municipality, the necessary amount of time has to be put in. *Nevertheless, after discussions with other municipal managers, they confirm they would not get headaches over this and that they would write down the same information than the previous year, with the exception of some information*" [comment b45 – *emphasis added*]

First, referring to the mandatory data transmission form as a 'survey' is a reminder that performance data are not audited in Quebec's municipal benchmarking system. In that regard, the values for the indicators are obtained though self-administered surveys. If we are to trust this public servant, some managers would simply not take the time to transmit accurate information about their municipal services. This idea that data is inaccurate was

also discussed, although in more diplomatic terms, by an Assistant-General Manager in a small municipality of less than 2000 residents.

> "(…) The service directors should also be made aware of collecting data correctly, which is not the case in our [Quebec's] small municipalities because the workload is too great for service directors and the General Manager. Thus, the management indicators are set aside, except for finances" [comment b46]

Another comment to that effect formulated by another General Manager in a small municipality of less than 2000 residents.

> "(…) For small municipalities, most of the time, the management indicators are prepared hastily with the little information that we have" [comment b52]

This adds to the idea that the quality of the data, let alone the external comparative data, is seen as leaving much to be desired.

Another insight from the qualitative analysis of the comments has to do with the most frequent theme in the general comment section: the lack of time. Here is the full quote from a General Manager-CFO working in a municipality of less than 2,000 residents.

> "When I started my functions in 2005, the management indicators seemed like a task to be accomplished by the GM/CFO. Information to collect and all and all. After discussion with my colleagues (ADMQ[xiii]) the necessity of the management indicators is not motivated. *For many of us, it is only two (2) hours spent on a software to complete a document* that will end up in the garbage when presented to council members" [comment b59 - *emphasis added*]

Given the annual basis of data transmission to MAMROT, the two hours this manager is referring to are two hours spent in a year. In contrast, here is another comment made by yet another General Manager of a municipality of less than 2,000 residents.

[xiii] A professional association of municipal directors of small municipalities

> "Without a software which provides comparatives, it is not very useful for us; *we work a lot with the budget and its bimonthly follow-up*" [comment b87 - *emphasis added*]

Although both pieces of information are from municipalities of similar sizes, it is unclear how typical the respective effort and time devoted to performance information and budgets are. However, the contrast between devoting two hours a year to performance information collecting and transmission, and working on budget follow-ups twice a month is evident.

All in all, the comments for the open-ended survey questions reveal that managers have many recriminations about the very idea of having their municipality being measured.

Focus groups of municipal actors on the use of management indicators

On January 28th, 2010, municipal managers were invited to meet in Saint-Hyacinthe, QC, for a day of activities devoted to the municipal management indicators. Among the activities were a number of presentations from managers, two presentations from academics (including myself), presentations of best practices in the use of indicators by managers, and a discussion session for participants. A total of 179 participants representing 100 municipalities, MRCs and municipal agencies were present at this day of activities organized by the Partners on Municipal Management Indicators Committee [*Comité des partenaires des indicateurs de gestion municipaux*], a coalition of professional associations and governmental agencies. Evidently, the participants were motivated enough to travel and devote a day to the indicators: they are a self-selected sample of managers. However, the participants were by no mean all convinced veteran users of performance indicators. Anecdotally, at the beginning of the day, one keynote

speaker asked the crowd to raise their hands if they considered themselves as users of performance information: roughly half raised their hands. The sample was one of convenience. This being said, it would have been next to impossible to organize focus groups with so many managers in such a vast province, if managers were selected by representative stratified samples.

The discussion session for participants consisted of focus groups discussing five selected findings on the use of performance information from the survey discussed earlier. The participants were divided in nineteen assigned tables of eight to ten participants. Participants were grouped according to the size of their municipalities. They were given for instructions that one person per table should take notes about the participants recommendations, but only after a consensus was reached. Participants were also asked to answer the five questions. They were asked to stick to the question and to forgo discussions about the relevancy of the indicators. This is important in understanding the focus group data, because the answers tend to be more proactive and less the work of nay-sayers as in the survey. Every fifteen to twenty minutes, the answer sheets were collected to make sure that each table would have enough time to answer each question. The whole focus groups activity lasted 1 ½ hour. The Partners on Municipal Management Indicators Committee agreed to release the focus group sheets to the author. Some of the answer sheets bear the number of the table of the participants. However, the data are truly anonymous, as the answers represent a group consensus of eight to ten people and the author does not have access to the roster of who were assigned to which table.

There were five questions for the participants to discuss. Each question was introduced by a finding from the survey presented earlier. The first question that was discussed by managers was the following:

1 Use and integration of management indicators in my municipality.
59% of users and 81% of non-users who answered the survey say they do not know how to include the management indicators in their decisions.
1.1 According to you, what are the actions and initiatives that could be put forward to foster use and inclusion of the municipal management indicators to introduce them in your decisions?

Recurring themes from the nineteen tables were demands for training about how to analyze performance indicators, and suggestions to create groups were analyses of indicators would be done in groups. There were also mentions to reduce the two-year delay for obtaining the data. For example, it was suggested by some participants that one step to include the management indicators in their decisions would be to

> "Shorten the transmission of indicators, so they become budgetary planning tool (available in June)" [suggestion from table #4]

Although other suggestions were made to be stricter with the transmission deadline of performance information, the suggestion above was one of the more specific. It squared with other suggestions about integrating the performance information into the budget cycle

> "Introduce in the budget process" [suggestion from an unidentified table]
>
> "Mandate the auditor to take a stand, to make an audit of the indicators; mandatory to report in the financial statements, because elected officials are not interested" [suggestion from table #19]
>
> "Presentation in the budget process" [suggestion from table #4]

Requests to move toward performance budgeting is surprising, given that the results from the survey suggest that even symbolic use is not widespread. However, they were only three suggestions in regard to performance budgeting out of the nineteen tables. Another theme about stepping up performance management that was present in the suggestions is the development of objectives or targets. Four focus groups reached a consensus that objectives/targets would help including the indicators into their decision making. One of these four tables that mentioned the development of objectives also wrote two items on sensemaking.

> "A good ratio, what is it?" [suggestion from an unidentified table]
>
> "Where are we [in comparisons]? Are we good are not?" [suggestion from an unidentified table]

Their suggestions following the development of objectives express their perception that one way to increase the use of performance information if to make the concerted effort to analyze the information. This idea complements the demand for training on performance indicators' analysis.

The second question on which participants had to express themselves was a follow-up question to Question 1.1.

1 Use and integration of management indicators in my municipality.
59% of users and 81% of non-users who answered the survey say they do not know how to include the management indicators in their decisions.
1.2 How could we make all managers include their indicators in their management style?

Question 1.2 is very similar to Question 1.1. Accordingly, answers for the second question were very similar to the ones for the first question. Demands for training were

even more frequent for Question 1.2. Some of the themes about delays in accessing indicators were once again present. Some suggestions were related to how management indicators are perceived. For example, one group of participants reached a consensus that in order to integrate the indicators in their management style, MAMROT should:

> "Start with one indicator (go step by step)" [suggestion from an unidentified table]

This remark is interesting, especially in the face of answers to Question 2, where managers commented on the shared perception that indicators paint an incomplete image of municipal performance. Two other suggestions were made on how to integrate indicators into management style. The shared theme is about making performance management look less threatening.

> "Explain to managers that it is a complementary tool to decision making" [suggestion from unidentified table]

> "Sell as [a] management [tool] and not as [a] performance [tool]" [suggestion from an unidentified table]

These suggestions were not shared by other focus groups. However, it is telling that two groups of managers found it necessary to express their concerns about how indicators are seen as judgmental and as a replacement of managers' discretion, given the non-intrusive non-threatening nature of Quebec's municipal benchmarking system. This will be discussed in the analysis section of this chapter.

Question 2 was about a topic that was raised in the comments from the survey that is the perception of indicators as a simplistic tool.

2. Portrait

70% of users and 86% of non-users mentioned in the survey that management indicators paint an incomplete image of municipal services

According to you, what could paint a complete image of all services?

Four main themes could be identified in the suggestions from the different focus groups. First, efforts to improve the imputation of cost accounting for the cost indicators were mentioned. Second, some tables mentioned that it would be impossible to have a complete image of all services. One comment to that effect is especially telling, as it can be linked with the fear of being measured. An asterisk was added to the comment of one focus group to the effect that:

> "*never a complete image, but it would be favorable to each municipality, but acceptable to all" [suggestion from table #9]

This suggestion is of interest in light of the third and fourth themes. The two following irreconcilable themes are about suggestion to add indicators, on one hand, or to replace mandatory indicators to let municipalities develop their own, on the other hand. What table #9 is saying in their comments is that indicators should not be unfavorable to any and be acceptable to all. Even if it did not square with the question at hand about the complete depiction of municipal services, one of the tables suggested that:

> "Train to explain how it is a tool and not a judgment" [suggestion from table #17]

For some managers, it would not be as much as showing a complete image of municipal services, as it is about making some painting the performance of some municipalities as being suboptimal. This brings us to Question 3.

Question 3 was there to foster discussion about the perception that management indicators would be misinterpreted or misunderstood.

3. Interpretation
82% of users and 93% of non-users mentioned in the survey that they fear that the management indicators would be misunderstood or misinterpreted.
What would enable you to avoid the misunderstanding or misinterpretation of results of the management indicators?

Once again, some themes that appeared earlier for Questions 1.1, 1.2 and 2 appeared for this question. Managers mentioned that data transmission delays should be shortened so that data can be recent. Once more, there are also suggestions to the effect that that cost accounting should be standardized, so that comparisons are not misunderstood.

A suggestion that was shared by some of the focus groups is to make sure to include the influential factors not only in reporting, but also in the online anonymous database of the password-protected portal accessible to managers. Three suggestions that set themselves apart from other themes have in common a very defensive stance vis-à-vis performance indicators. On the misunderstanding or misinterpretation of results of the management indicators, groups of managers expressed that it could be avoided:

> "Not using the indicators to create competition" [suggestion from table #7]

Another focus group reached consensus with a very similar idea.

> "Explain and "do not judge" [suggestion from table #17]

One comment is even more telling variant of the no competition/no judgment theme.

> "See if MAMROT could analyze global data and notify cities that are outside the average so they could verify their numbers, identify the influential factors" [suggestion from table #19]

Verifications of data could be sought. What is interesting here, what is implied is that the values of indicators are due to errors of calculations and exogenous factors influencing the data. It is as if it is unconceivable that some municipalities stray away from the mean value of an indicator because some services are performed more or less efficiently. This is linked to the fifth and last question asked to participants.

Question 4 was about opinion that comparisons between municipalities are impossible and should not be done.

4. Comparison Among the obstacles to the use of management indicators, 67% of users and 81% of non-users mentioned in the survey, identified the impossibility to compare their municipality's results to other comparable municipalities. Other than MAMROT's software, what are the means, according to you, which could help you compare your results to those of similar municipalities.

The suggestions to this question were a lot less defensive than those of previous questions. Themes encountered before were reproduced here; among them are suggestion to the effect of reducing delays in obtaining recent data, and making sure that cost indicators are calculated the same way. There were also expressed demands to create informal groups of municipalities, akin to benchmarking clubs, were data would be exchanges, analyzed and discussed jointly. There were also many demands for MAMROT to change the online database in the password-protected portal to enable queries that are either based on something else than size, or even cases where anonymous queries would be set aside for more flexible ones. This would mean a comparative tool not unlike Ontario's MIDAS tool.

Unlike the suggestions for Question 1.1, 1.2, 2, and 3, there were no suggestion from Question 4 that let transpired a malaise or a defensive position toward the municipal performance indicators.

Discussion of the qualitative data

At first, it might be surprising to witness this level of opposition to management indicators. The indicators started in 1999, with consultations with a myriad of municipal associations and associations of managers, CFOs, and accountants. Pilot projects about the indicators started between May 2001 and May 2002. Data collection for the 19 mandatory indicators started for every Quebec municipality in 2003. Since 2005, it has been mandatory for municipalities to make the data public. Discussing the indicators during a council meeting fulfills the mandate of 'making the data public.' In 2007, after consultations with professional associations, some indicators were dropped. Furthermore, there are no consequences for municipalities from MAMROT, no matter the value of the indicators. MAMROT does not even specify what constitute good, bad, better or worse performance. In the current electronic portal where all the data are gathered, managers need a password to access comparative data. Even there, in the password protected electronic tool, comparative data is aggregated and anonymous. The public is left out. Compared to other mandatory municipal benchmarking initiatives in the provinces of Ontario and Nova-Scotia; and in Wales, Scotland, Ireland, Britain, Norway, and New South Wales, Australia, it is by far the most opaque, lenient and non-threatening. Proponents of performance measurement can wonder on what grounds a sizeable

proportion of municipal managers in Quebec feel hostility to such an undemanding system that was implemented gradually with ample participation from practitioners.

One of the distinctive characteristics of municipalities in Quebec is that there are many very small municipalities. In these very small municipalities, there are often only a handful of administrative employees. Sometime, the entire bureaucracy in these municipalities is a single employee, who might not even work full time. Management indicators are seen as an addition to their task. The comments about the inapplicability of performance indicators in small municipalities should not be dismissed out of hand. The perceptions of many municipal managers in small rural municipalities are that indicators are burdensome and time consuming: they would be impractical. Recurrent themes about the complexity of performance indicators, the burdensome and heavy nature of using performance indicators are frequent. This seems to be the perception. Once this is said, one has to ponder if using internal and external comparative data is really more complex than an annual budget. The lack of time, which is another frequent perceived impediment to use performance information, is questionable. Comparative performance information has been packaged by MAMROT (and formally by CPEGM) for years. On the other hand, managers have to prepare the annual budget themselves, often months in advance. On site observational research in municipalities would be needed to inform us about the complexity and the actual time devoted to the use of performance information and the preparation of the annual budget.

The perception of managers that uncooperative elected officials are impediments to the use of performance indicators is not new. A recent study by Taylor (2009) in Australia had similar findings. Questionnaires and semi-structured interviews of twenty-four high-ranking state officials throughout Australia revealed that managers perceived that politicians were less committed to performance measurement than managers would like. Some of the quotes from her interviews are telling:

> "I don't know that the politicians take them [PIs] seriously. I can see how they are a good mechanism for accountability, but I am not convinced that politicians see it that way. . . we try to inform politicians and all the rest of it, but I think they've got other agendas, unfortunately . . . . Their agenda is to get re-elected and to stay in office and all that." (Taylor, 2009:862)

> "It is fair to say that judgments and decisions made by politicians, by their very nature are politically focused, and it would take an exceptional piece of performance information to turn them around." (Taylor, 2009:862)

The comments are from state officials instead of municipal officials; they refer to Ministers and not to Mayors and council Members; and originate from the other side of the globe. Nevertheless, they reverberate with the experience of local managers in Quebec. Furthermore, most of the sixteen recurring themes from table 38, with the exception of the theme about small municipalities, can be equated with the original list of twelve barriers by Siverbo and Johansson (2006:283-284), from which the list of fourteen barriers in the survey is derived. The possibility that the list of barriers from question 7 of the survey influence the comment section of question 1 cannot be methodologically ruled out. However, two elements can suggest a different reading of the findings. First, there are overlaps between previous lists of barriers or obstacles to greater use of performance (Ammons, 1985:295; Poister and Streib, 1999:332; McAdam and O'Neill, 2002:452-453;

Johnsen, 2005:11; Pollanen, 2005:17; Rantanen *et al.*, 2007:428) and the comments of municipal managers in Quebec. Second, there are similarities between responses from Swedish municipal managers, comments from top state managers in Australia and the comments of municipal managers in Quebec regarding the barriers or obstacles to greater use of performance. Internationally, managers who operate in different performance management systems with different characteristics express similar reasons to explain why they do not use performance information more. There might be constants inherent in performance information why public managers do not feel compelled to use information they are suggested or ordered to collect, compile, transmit and report.

The theme about the impossibility of comparing municipal services can be at first surprising within a benchmarking project. This is at odds with the official position by MAMROT, which suggest that comparisons should be made. An example of this idea means can be appreciated for water treatment. Water quality in Quebec is bounded by the *Règlement sur la qualité de l'eau potable* [Ruling on the quality of potable water] (c. Q-2, r.18.1.1) under the *Loi sur la qualité de l'environnement* [Law on Environmental Quality] (L.R.Q., c. Q-2, a. 31, 45, 45.2, 46, 87, 109.1 & 124.1). It imposes strict guidelines on many aspects of water treatment. Among them are: operator's training; bacteriological, physic-chemical controls and norms; specification of water trucks; minimal numbers of water sampling modulated by population size; standardized technical guides for articles 53 and 53.0.1 implementation; and standardized weekly data transmission forms. Despite all of these guidelines, there would be in Quebec for the 925 municipalities (out of 1113) who do offer a water treatment service, 925 different and irreconcilable sets of goals

about the quality of water and the cost of its treatment could not be compared in any ways. The same would hold true on other municipal services like snow plowing and street repairs. More qualitative research with open-ended questions about operational goals would be needed to assess if there are similarities and recurring themes in objectives that are pursued by municipalities. Additional studies on the presence of comparisons between municipalities on managers' remuneration would also be needed to see the proportion of managers that are consequent about their 'no comparisons are possible' stance.

Lastly, managers who made comments mentioned that they do not have the time to work with management indicators. Many complained that collecting performance information necessary for the transmission on the mandatory indicators was too time consuming. Two different managers from small municipalities of less than 2,000 residents revealed more about the efforts and times that were devoted to performance indicators and the budget. In one case, a General Manager wrote that he/she (and other counterparts) would spend two hours a year to collect and transmit the information on the indicators to MAMROT. The transmission form was referred to as a 'survey'. This was not the only time the form was referred as such. In another case, a General Manager explained that bimonthly follow-ups were done for the budget. This contrast in devoted time and energy, both from managers who admittedly do not use performance indicators, is manifest. This put the frequent 'lack of time' theme in a different light. Two hours to devote to a 'survey' is indeed a lot of time. I leave to the reader the judgment about the validity of the 'lack of

time' argument, when two hours a year are devoted to a management tool that is supposed to complement other management tools.

After reading the suggestions to the focus groups and the comments from survey takers, it is clear that there are managers who would like to see more indicators, more current indicators. They also mentioned that they would like to receive help to analyze their results and want training to become better at it. On the other hand, there are managers who either would like the management indicators to go away, or they would like them to be devoid of external comparisons or any form of evaluation. If analyses would be done, it should be in such a way that none of the 1113 municipalities are identified as offering one service that is deemed suboptimal. This, of course, flies in the face of the spirit behind performance measurement, where comparisons are sought to foster transparency or performance improvement.

The qualitative analyses of survey comments and focus groups suggestions add to the quantitative analyses. What comes out of the quantitative analyses is that management indicators are not widely used or only used occasionally. What comes out of the qualitative analyses are that many managers do not find the indicators relevant. Many others are scared of being evaluated, even if Quebec's benchmarking system was designed and is run to minimize that fear.

Chapter 6: Conclusion

Summary

A copious literature addresses the questions of which agencies would implement a performance measurement initiative, the benefits and limits of this tool, and the best practices for data collection. There is now an emerging literature on the use of performance information. The present research adds to the performance information use literature by studying a systematic mandatory performance measurement system in a North American context. Additionally, this is the first piece of research we are aware of that takes into account the performance levels as a factor explaining the use of performance information. Also to our knowledge, this is the first study that tries to understand how managers in the public sector make sense of performance information.

In the performance measurement literature, what happened once data for performance indicators are collected is not well understood. Knowing the management, budgeting and reporting functions for which performance information is reportedly used, matters for provincial officials in Quebec. After all, performance measurement is an information-based managerial tool: if it is not used, the efforts and resources that have been invested since the first public consultations in 1999 could have been invested elsewhere.

This research tried to answer four questions. The questions are:

R1. Which factors account for the uses of performance measurement by municipal managers?

R2. What are the comparison levels being used by municipal managers in interpreting performance measures?

R3. How are targets set by municipal managers?

R4. While interpreting their own performance, when do municipal managers deem that the municipality's performance is (a) a failure, (b) unsatisfying, (c) satisfying, (d) a success?

The answer to the first research questions is that there are three factors accounting for the use of performance measurement by municipal managers. The first factor is related to the unwillingness of managers to use performance indicators. It could not be demonstrated that the perceived inability to use indicators and the perception of being prevented from using the indicators, actually influence the different uses of performance information. The second factor influencing the use of performance information is the leadership of managers in performance management. No correlations were found between the various uses of performance information and other internal characteristics like clear priorities, links between priorities and community needs, internal partnership working, and political leadership. Contrary to the claims of managers found in chapter 5, no link was established between the rural character or the size of a municipality, and the use of performance information. The third factor influencing the use of performance information is performance itself. Municipalities that perform better than other municipalities of comparable size tend to use performance indicators to a greater extent.

The answer to the second question is that General Managers choose to ignore available data external to their municipality and compare their current performance to their past performance.

There is not a satisfactory answer to the third question. To put it bluntly, for the most part, no targets are set.

The answer to the fourth question is complex. Negative performance is almost never considered a failure: it is mostly deemed unsatisfactory, although some think it can be satisfactory. Neutral performance is overwhelmingly seen as satisfactory. Positive performance is seen as either a success or satisfactory, almost in equal measures. Managers think more in terms of satisfactory and unsatisfactory than in terms of success and failure. They think more in positive than in negative terms. Managers do not think of symmetrical performance in symmetrical terms. In the view of managers, achieving positive performance is seen much more positively than negative performance is seen negatively. For between one out of four and one out of three General Managers in Quebec, none of the twelve scenarios could be interpreted in a negative way.

Implications for Quebec Municipal Governance

After an initially timid set of objectives, came successive scaling backs of objectives. Currently, MAMROT mandates municipalities to collect performance information. It is not clear to what end. As we have seen in chapter 2, Quebec's benchmarking system is not tied with provincial transfer, there are no required targets, and there are no obligations other than transmitting the data to MAMROT and presenting it at meetings of municipal council.

Data in Quebec's systems are not current like the mandatory municipal systems in Nova Scotia, Canada, Wales, Scotland or Norway. The performance measurement system in Quebec is not accessible to the public like the systems in Nova Scotia, Canada, New South Wales, Australia, Wales, Scotland, England or Norway. The Municipal Management Indicators system in Quebec relies on managers to analyze their performance and take remedial actions, if need be. To achieve that, managers, shielded from public scrutiny, can compare their performance using two-year old comparative data. Managers cannot compare themselves to specific municipalities, like in the system of neighboring Ontario; they can only compare themselves to quartile data of anonymous amalgamation of municipalities.

In Quebec, there are no targets like in England. Targets, especially if they are set with external comparison information, would be one of the most meaningful referent points for either performance improvement or transparency. In Quebec's Municipal Management Indicators system, targets are voluntary. Since Fall of 2009, they are not part of MAMROT's official objectives for the benchmarking systems. What's more, since 2003, very few managers set targets for their indicators. When targets are set, they are set internally, not with external information.

In Quebec, the data are not audited like they are in New South Wales, Australia, Wales, Scotland, England or Norway. One of the recurrent themes in the survey comments and the focus groups is that the values of the indicators are not seen as trustworthy. Managers think that the data, especially for cost indicators, might be inaccurate. Managers are

aware of that because they know they could transmit inaccurate data themselves (see McAdam and O'Neill, 2002:452-453). If MAMROT wants managers to use the system, managers have to consider the system to offer quality data. To achieve that, municipal managers should not be allowed to transmit inaccurate data to the system and get away with it. One other management tool, the budget, has a safeguard against this: there are called auditors.

One of the four objectives that remain (from the original eight) from MAMROT's list, is that the performance of municipal services would be improved. Currently, on the password-protected portal, managers can compare their performance with anonymous aggregates. Depending on indicators, the comparisons that are available with referent points are comparisons with municipalities of comparable size or characteristics. MAMROT does not stipulate what constitutes good and bad performance. It does not even suggest what constitutes better or worse performance. Anonymous comparisons are broken down by quartiles. However, the quartiles are not indicative of better or worse performance: they are solely organized by numeric values. How would MAMROT know if performance improved? How would managers know? In the absence of official guidelines of what constitutes a successful/failure or satisfying/unsatisfying performance, managers' perception of sensemaking is key. The role of performance information is to inform managers. Actions do not stem from collecting performance indicators; actions stem from the analysis of performance indicators. Actions stem from the realization that, after analysis, the organization failed to perform the way it could or should. According to

most of the General Managers in Quebec, failure does not exist. According to many of them, unsatisfactory levels of performance do not exist.

Comparisons with targets and with other municipalities would inform an analyst about what the values of indicators mean. Data need to be used before they are analyzed. In Quebec, most municipalities do not use the indicators. Only the rarest of municipalities set targets for themselves. Most municipalities that do use the indicators do not compare themselves to others. A common comment from managers that do not use the indicators is that the very idea of comparing municipalities among themselves is preposterous. One of the four remaining objectives of MAMROT is external comparisons.

Quebec's mandatory municipal benchmarking system shares a lot of characteristics with Ontario's system: many cost indicators, long delays in obtaining and transmitting data, results presented in quartiles, results hidden from the public. In her dissertation on performance reporting in Ontario, Schatteman (2009:136) found that "(…) public managers certainly have negative perceptions of the public performance system and do not use it to a great extent in their decision-making." Schatteman (2009:142) concludes that managers in Ontario only go through the motions, they are not invested. The same could be said about the situation in Quebec. MAMROT asks municipalities little more than data be collected and transmitted to the provincial capital. A majority of municipalities do just that.

Policy Recommendations

Recommendations only make sense in relation to objectives. MAMROT's objectives were scaled back with time. The objective of using municipal management indicators as a way to be accountable to citizens was dropped sometime between 2004 and 2008 (MAMSL, 2004; MAMR, 2008). MAMROT's objective of fostering the understanding of service quality and financial health of municipalities was abandoned between 2008 and 2009 (MAMR, 2009). The objectives of using indicators to foster planning and target setting, and preserve resources were also discarded during that period.

General recommendations

It is widely recognized that performance or management indicators have two main benefits: (1) increased transparency, and (2) potentially improved performance. It is usually understood that these benefits are mutually exclusive. MAMROT's current list of objectives suggests that the performance improvement route is not being pursued. Thus, we recommend that:

Either,

- Reintroduce the objective of "Plan municipal services better and determine targets with a better understanding of the situation" as it was in July 2009.
    - Currently, depending on the indicators, only 1,14% to 8,52% of surveyed managers say that they established a target for their services. This is relevant, as research showed that municipalities that set targets perform better than municipalities that do not (*for example*, Revelli and Tovmo, 2007:131).

  - Is has been proven that a voluntary system, in North Carolina (U.S.A.) where information is collected for general information purposes, does not foster performance improvement in municipal services (Williams, 2005)

or

- In relation to transparency, consider moving away from the model in Ontario, Canada, and emulate models in Nova-Scotia, Canada, New South Wales, Australia, Wales, Scotland, England, and Norway: make the data available to the public.

More generally,

- Continue the practice of including stakeholders in the iterations of the Municipal Management Indicators

Types of indicators

- Continue on MAMROT's current trend to limit the number of cost indicators and concentrate on indicators of service effectiveness.
- Make some of the currently facultative indicators (libraries, fire services, etc.) mandatory.
- Add mandatory indicators for police services.

Data collection

- Shorten the deadline between data collection and data transmission.
  - In date of late March 2010, Quebec's benchmarking password-protected portal just uploaded the 2008 comparative data. Other mandatory systems in Ontario, Canada and New South Wales, Australia, also have relatively outdated data in their systems. Mandatory benchmarking systems in Nova Scotia, Canada, Wales, Scotland, England and Norway have 2009 data.

- Require that municipalities, if they decide to contract out on municipal services that are being measured, still have to collect and report the mandatory information to MAMROT.
- Have the data of indicators audited, as it is the case of more trusted managerial data contained in budgets and budget documents.
    - This would meet one of the most frequent demands formulated by managers in relation to the mistrust of comparative data

Functionality of the internet portal

- Allow managers accessing the password-protected web portal to perform ad-hoc comparative queries.

Analysis

- Assist managers by offering analytical support of indicators' data, at the MRC-level or at the provincial level.
- Validate influential factors, so they could be used to set up comparative groups that take more characteristics into account beside a municipality's population size.
    - Follow up a 2004 recommendation about using validating influential factors to categorize municipalities (Guindon and Bellavance, 2004).
- Study the possibility to include more sophisticated analyses of indicators results that take into considerations the resources and service levels for all municipal services simultaneously (for example, Spottiswoode, 2000).

Future Research

Following are two suggestions for further research:

First, on the use of performance information, future research should find a way to reduce dependence on perception data on the part of managers. Observing the direct *real* use of performance information would enhance our knowledge of performance measurement. Quebec's benchmarking system offers such an opportunity to observe the *real* use of comparative performance information. Within the password-protected web-portal where comparative information is available for municipal managers in Quebec, there is an electronic log of who actually consulted the information. This would provide a more reliable variable for the use of performance information.

Second, follow-up interviews should be conducted about the sensemaking results from the survey. Chapter 5, the qualitative chapter from this study, offered precious insights into the use of performance measurement. A qualitative assessment of the quantitative findings on sensemaking would facilitate deeper knowledge of this emerging issue. It would be edifying to hear directly from managers the reasons why they shy away from negative analyses of negative performance. It would also be informative to know organically how managers make sense of their data, free from a predetermined survey question.

Broader Implications

The difference between performance measurement and performance management is that in the latter, performance information is included in decision-making and operations. In the modern era, benchmarking in the public sector was borrowed from the private sector. To maximize the chances that it will be implemented, and to meet its purposes, a benchmarking system should recognize the specificity of organizations in the public sector. A private organization exists to foster the private interests of few individuals. A public organization exists to foster the public interest of a collectivity. The notions for accountability and performance are different.

Private organizations are accountable to their stakeholders. If they engage in benchmarking activities, they do not have to report the comparative results to non-stakeholders. Non-stakeholders do not have a vested interest in their performance. Public organizations are accountable to the public. The public is stakeholders. The public has a vested interest in public sector organizations' performance. Denying this accountability does not serve the public interest.

Private organizations operate in a competitive market. Although their bottom line, profit, is partially dependent on competitive forces, it is affected by these forces. Other standard performance measures like market shares are determined by competitive forces. It is normal that the performance measures are designed in competitive ways. Private organizations have flexibility to change their services entirely, add new products, eliminate some services or select their customers. Public organizations do not have this

flexibility. Public organizations do not operate in a competitive market. They do not have a bottom line. Assessing how they are doing necessitates comparing what they are doing with what they were mandated to do. What they are mandated to do is operationalized as targets. Since they have no bottom line, the other way to evaluate a public agency is to see how other agencies with the same mandate are doing. Different public agencies pursuing the same mandate in different locations are not in competition with each other. There are no market shares: it is not a zero-summed game. It is not normal that the performance measures are designed in competitive ways. Organizing comparative performance in quartiles means that the performance measures are designed in a competitive way. The only way for a municipality to get into the top quartile is for another municipality to get out of that quartile. Quartiles are a zero-summed game.

An improvement from comparative quartiles would have many attributes. It would take into account constraints facing public organizations, like minimal levels of service delivery and governmental norms. It would take into factor in certain geographical differences, like terrain. It would take into account that to assess the performance of a municipality, all services should be evaluated jointly. It would take into account that budgetary choices are made and that different municipal services are prioritized in different municipalities. It would take into account that there is not one best way to manage a municipality: many models exist. It should take into account that municipalities, are not (strictly speaking) in competition with each other: it is not a zero-summed game. It should recognize and reflect that municipalities that are improving their performance, even if others are improving their performance more, deserve credit. In

short, the comparative design of a benchmarking system in the public sector should recognize complexity. Quartiles do not recognize complexity.

Some tools can meet some of the items just enumerated. There is one tool that has the potential to meet all of these items. This tool was presented in chapter 1: data envelopment analysis. It is not used very often by the public agencies themselves (with the notable exception of Finnish municipalities, see Johnsen and Vakkuri, 2006:296). It is used often by academics evaluating public agencies. This tool recognizes complexity.

By design, performance measurement simplifies reality. There is a balance to be struck between the complexity involved in managing a public sector organization and the tool used to understand this reality. Unless managing a public organization becomes less complex in the foreseeable future, performance measurement should espouse complexity.

BIBLIOGRAPHY

Aaron, W. (2008). "What Gets Measured Sometimes Gets Managed." Government Finance Review **24**(4): 24-28.

Afonso, A. and S. Fernandes (2008). "Assessing and Explaining the Relative Efficiency of Local Government." Journal of Socio-Economics **37**: 1946–1979.

Alden, J. (2006). "Measuring the "Unmeasurable"." Performance Improvement **45**(5): 7-11.

Alexander, J. A., B. J. Weiner, *et al.* (2006). "Quality Improvement and Hospital Financial Performance." Journal of Organizational Behavior **27**(7): 1003-1029.

Allison, G.T. (1992). "Public and Private Management: Are They Fundamentally Alike in All Unimportant Respects?" in R.J. Stillman III, ed. Public Administration: Concepts and Cases, Fifth Edition, Boston, MA: Houghton Miffin Compagny: 282-288.

Ammons, D. N. (1985). "Common Barriers to Productivity Improvement in Local Government." Public Productivity Review **9**(4): 293-310.

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

Ammons, D. N. (1999). "A Proper Mentality for Benchmarking." Public Administration Review **59**(2): 105-109.

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.

Ammons, D. N. and W. 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.

Andrews, R., G. A. Boyne, *et al.* (2005). "External Constraints on Local Service Standards: The Case of Comprehensive Performance Assessment in English Local Government." Public Administration **83**(3): 639–656.

Askim, J. (2004). "Performance Management and Organizational Intelligence: Adapting the Balanced Scorecard in Larvik Municipality." International Public Management Journal **7**(3): 415-438.

Askim, J. (2007). "How Do Politicians Use Performance Information? An Analysis of the Norwegian Local Government Experience." International Review of Administrative Sciences **73**(3): 453–472.

Askim, J. (2008). Determinants of Performance Information Utilization in Political Decision Making. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 125-139.

Askim, J. (2009). "The Demand Side of Performance Measurement: Explaining Councillors' Utilization of Performance Information in Policymaking." International Public Management Journal **12**(1): 24-47.

Askim, J., Å. Johnsen, *et al.* (2008). "Factors behind Organizational Learning from Benchmarking: Experiences from Norwegian Municipal Benchmarking Networks." Journal of Public Administration Research and Theory **18**(2): 297–320.

Bailey, T. C. and P. J. Hewson (2004). "Simultaneous Modelling of Multiple Traffic Safety Performance Indicators by Using a Multivariate Generalized Linear Mixed Model." Journal of the Royal Statistical Society (Series A) **167**(3): 501-517.

Balk, W. L. (1993). "Democratic Action and Productivity Improvement." Public Productivity & Management Review **16**(4): 437-445.

Ball, A., M. Bowerman, *et al.* (2000). "Benchmarking in Local Government under a Central Government Agenda." Benchmarking: An International Journal **7**(1): 20-34.

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

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

Behn, R. D. (2008). 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.

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

Berman, B. J. C. (2008). "Involving the Public in Measuring and Reporting Local Government Performance." National Civic Review **97**(1): 3-10.

Berman, E. and X. Wang (2000). "Performance Measurement is U.S. Counties: Capacity for Reform." Public Administration Review **60**(5): 409-420.

Bernstein, D. (1999). "Comments on Perrin's Effective Use and Misuse of Performance Measurement." American Journal of Evaluation 20(1): 85-93.

Bevan, G. and C. Hood (2005). What's Measured is What Matters: Targets and Gaming in the English Public Health Care System. The Public Services Programme: Quality, Performance & Delivery. London, UK**:** 1-46.

Bevan, G. and C. Hood (2006). "Have Targets Improved Performance in the English NHS?" British Medical Journal **332**: 419-422.

Bird, S. M., D. Cox, *et al.* (2005). "Performance Indicators: Good, Bad, and Ugly." Journal of the Royal Statistical Society (Series A) **168**(1): 1-27.

Bititci, U. S., T. Turner, *et al.* (2000). "Dynamics of Performance Measurement Systems." International Journal of Operations & Production Management **20**(6): 692-704.

Bolton, M. (2003). "Public Sector Performance Measurement: Delivering Greater Accountability." Work Study **52**(1): 20-24.

Bouckaert, G. and W. Balk (1991). "Public Productivity Measurement: Diseases and Cures." Public Productivity & Management Review **15**(2): 229-235.

Bovaird, T. (1996). The Political Economy of Performance Measurement. Organizational Performance and Measurement in the Public Sector: Toward Service, Effort and Accomplishment Reporting. A. Halachmi, and G. Bouckaert. Westport, CT, Quorum Books: 145-165.

Bovens, M., P. t. Hart, *et al.* (2001). The State of Governance in Six European States. Success and Failure in Public Governance. M. Bovens, P. t. Hart and B. G. Peters. Cheltenham, UK, Edward Elgar Publishing**:** 641-658.

Bowerman, M., A. Ball, *et al.* (2001). "Benchmarking as a Tool for the Modernisation of Local Government." Financial Accountability & Management **17**(4): 321-329.

Bowerman, M., G. Francis, *et al.* (2002). "The Evolution of Benchmarking in UK Local Authorities." Benchmarking **9**(5): 429-449.

Boyne, G. (1997). "Comparing the Performance of Local Authorities: An Evaluation of the Audit Commission Indicators." Local Government Studies **23**(4):7-43.

Boyne, G. and G. Enticott (2004). "Are the 'Poor' Different? The Internal Characteristics of Local Authorities in the Five Comprehensive Performance Assessment Groups." Public Money & Management **24**(1): 11-18.

Boyne, G., J. Gould–Williams, *et al.* (2002). "Plans, Performance Information and Accountability: The Case of Best Value." Public Administration **80**(4): 691-710.

Boyne, G. A. (2006). "Strategies for Public Sector Turnaround: Lessons from the Private Sector?" Administration & Society **38**(3): 365-388.

Boyne, G. A., K. J. Meier, *et al.* (2006). Public Management and Organizational Performance: An Agenda for Research. Public Service Performance: Perspectives on Measurement and Management. G. A. Boyne, K. J. Meier, L. J. J. O'Toole and R. M. Walker. New York, NY, Cambridge University Press**:** 295-311.

Boyne, G. A. and A. A. Chen (2007). "Performance Targets and Public Service Improvement." Journal of Public Administration Research and Theory **17**(3): 455-477.

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

Braadbaart, O. and B. Yusnandarshah (2008). "Public Sector Benchmarking: A Survey of Scientific Articles, 1990–2005." International Review of Administrative Sciences **74**(3): 421-433.

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

Brusca, I. and V. Montesinos (2005). "Are Citizens Significant Users of Government Financial Information?" Public Money & Management. **26**(4): 205-209.

Butterworth, P., Gray, R.M. and J. Haslam (1989). "The Local Authority Annual Report in the UK: An Exploratory Study of Accounting Communication and Democracy." Financial Accountability and Management. **5**(2): 73-87.

Cameron, K.A., R.I. Sutton, *et al.* (1988). Readings in Organizational Decline: Frameworks, Research, and Prescriptions. Cambridge, MA, Ballinger.

Camp, R. (1989a). "Benchmarking: The Search for Best Practices that Lead to Superior Performance, Part I", Quality Progress **22**(1): 61-68.

Camp, R. (1989b). "Benchmarking: The Search for Best Practices that Lead to Superior Performance, Part II", Quality Progress **22** (2): 70-75.

Carmona, S. and A. Grönlund (2003). "Measures vs Actions: The Balanced Scorecard in Swedish Law Enforcement." International Journal of Operations & Production Management **23**(12): 1475-1496.

Carrington, R., N. Puthucheary, *et al.* (1997). "Performance Measurement in Government Service Provision: The Case of Police Services in New South Wales." Journal of Productivity Analysis **8**(4): 415-430.

Carvalho, J., M. Fernandes, et al. (2006). "Measuring Fire Service Performance: A Comparative Study." International Journal of Public Sector Management **19**(2): 165-179.

Castelli, A., D. Dawson, et al. (2007). "A New Approach to Measuring Health System Output and Productivity." National Institute Economic Review **200**(1): 105-117.

Catasùs, B., S. Ersson, et al. (2007). "What Gets Measured Gets...On Indicating, Mobilizing and Acting." Accounting, Auditing & Accountability **20**(4): 505-521.

Centre de promotion de l'excellence en gestion municipale (2005a). Guide sur la reddition de comptes des indicateurs de gestion municipaux auprès du citoyen. Québec, QC, Corporation des officiers municipaux agréés du Québec**:** 40p.

Centre de promotion de l'excellence en gestion municipale (2005b). Rapport concernant les travaux sur la reddition de comptes et l'optimisation des indicateurs de gestion municipaux. Québec, QC, Corporation des officiers municipaux agréés du Québec**:** 80p.

Chalos, P. and J. 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.

Chang, R., and K. Kelly (1994). Improving through Benchmarking: A Practical Guide to Achieving Peak Process Performance, Irvine, CA, Richard Chang Associates.

Charbonneau, É., F. Bellavance, and M. Holzer (2010). Are Reporting Practices Determined by Performance Levels?: Evidence from Quebec's Municipal Management Indicators System, Working paper. Montréal, QC, Centre de promotion de l'excellence en gestion municipal: 13p.

Clarkson, P. and D. Challis (2006). "Performance Measurement in Social Care: A Comparison of Efficiency Measurement Methods." Social Policy & Society **5**(4): 461–477.

Chapman, D. W. and J. R. Lombard (2006). "Determinants of Neighborhood Satisfaction in Fee-Based Gated and Nongated Communities." Urban Affairs Review **41**(6): 769-799.

Chrystal, K.A. and P. Mizen (2003). Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy, in P. Mizen (ed.) Central Banking, Monetary Theory and Practice: Essays in Honour of Charles Goodhart. Vol. 1. Cheltenham, UK, Edward Elgar: 221-243.

Chung, Y. (2005). Factors Affecting Uses and Impacts of Performance Measures in Mid-Sized U.S. Cities. Department of Political Science. Knoxville, Tennessee, University of Tennessee**:** 1-192.

Clarkson, P. and D. Challis (2006). "Performance Measurement in Social Care: A Comparison of Efficiency Measurement Methods." Social Policy & Society **5**(4): 461–477.

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

Coe, C. K. (2003). "A Report Cards on Report Cards." Public Performance & Management Review **27**(2): 53-76.

Cooper, W.W., Seiford, L.M. and K. Tone (2006). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Second edition. New York, NY: Springer.

Cordero-Ferrera, J. M., F. Pedraja-Chaparro, et al. (2008). "Measuring Efficiency in Education: An Analysis of Different Approaches for Incorporating Non-Discretionary Inputs." Applied Economics **40**(10): 1323-1339.

Cox, A. and I. Thompson (1998). "On the Appropriateness of Benchmarking." Journal of General Management **23**(4):1-20.

Coy, D. and K. Dixon (2004). "The Public Accountability Index: Crafting a Parametric Disclosure Index for Annual Reports." British Accounting Review **36**(1):79-106.

Courty, P., C. Heinrich, et al. (2005). "Setting the Standard in Performance Measurement Systems." International Public Management Journal **8**(3): 321-347.

Daft, R. L. and K. E. Weick (1984). "Toward a Model of Organizations as Interpretation Systems." Academy of Management Review **9**(2): 284-295.

Davies, M. (2004). "Performance Measurement in the UK Public Sector: Understanding Performance Indicators." Journal of Finance and Management in Public Services **3**(2): 31-47.

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): 260-270.

De Bruijin, H. (2002). "Performance Measurement in the Public Sector: Strategies to Cope with the Risks of Performance Measurement." International Journal of Public Sector Management **15**(6/7): 578-594.

de Lancer-Julnes, P. (2006). "Performance Measurement: An Effective Tool for Government Accountability? The Debate Goes On." Evaluation **12**(2): 219-235.

de Lancer-Julnes, P. and M. Holzer (2001). "Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation." Public Administration Review **61**(6): 693-708.

Diez-Ticio, A. and M. J. Mancebón (2002). "The Efficiency of the Spanish Police Service: An Application of the Multiactivity DEA Model." Applied Economics **34**(3): 351-362.

DiIulio, J. J. J. (1993). Measuring Performance When There Is No Bottom Line. Performance Measures for the Criminal Justice System: Discussion Papers from the BJS-Princeton Project. J. Q. Wilson. Washington DC, DIANE Publishing Company**:** 143-156.

Downe, J., C. Grace, et al. (2007). Comparing For Improvement: Local Government Performance Regimes in England, Scotland and Wales. The Public Services Programme: Quality, Performance & Delivery. London, UK**:** 1-25.

Drake, L. and R. Simper (2001). "The Economic Evaluation of Policing Activity: An Application of a Hybrid Methodology." European Journal of Law and Economics **12**(3): 173–192.

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.

Drake, L. and R. Simper (2003). "The Measurement of English and Welsh Police Force Efficiency: A Comparison of Distance Function Models." European Journal of Operational Research **147**(1): 165–186.

Drew, S. A. W. (1997). "From Knowledge to Action: the Impact of Benchmarking on Organizational Performance." Long Range Planning **30**(3): 427-441.

Duffy, J. A. M., J. A. Fitzsimmons, et al. (2006). "Identifying and Studying "Best-Performing" ervices; An Application of DEA to Long-Term Care." Benchmarking: An International Journal **13**(3): 232-251.

Dutton, J. E. (1993). "Interpretations on Automatic: A Different View of Strategic Issue Diagnosis." Journal of Management Studies **30**(3): 339-357.

Edwards, D. and J. C. Thomas (2005). "Developing a Municipal Performance-Measurement System: Reflections on the Atlanta Dashboard." Public Administration Review **65**(3): 369-376.

Eitel, D. F. (2004). "The Dynamics of Chronic Failure: A Longitudinal Study." Public Money & Management **24**(4): 243-250.

Ellig, J. (2007). "Scoring Government Performance Reports." Public Manager **36**(2): 3-8.

Enticott, G., R.M. Walker, G.A. Boyne, S.Martin and R. Ashworth (2002). Best Value in English Local Government: Summary Results from the Census of Local Authorities in 2001. Cardiff, Wales, Local and Regional Government Research Unit: 55p.

Farrell, M.J. (1957). "The Measurement of Productive Efficiency." Journal of the Royal Statistical Society Series A **120**(3): 253–281.

Feit, D. (2003). "Measuring Performance in the Public Sector." Cost Management **17**(2): 39-45.

Feldman, M. S. and J. G. March (1981). "Information in Organizations as Signal and Symbol" Administrative Science Quarterly **26**(2): 171-186.

Fiegenbaum, A. and S. Hart (1996). "Strategic Reference Point Theory." Strategic Management Journal **17**(3): 219-235.

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

Fountain, J. (1997). "Are State and Local Governments Using Performance Measures?" PA Times **20**(1).

Frederickson, D. G. and H. G. Frederickson (2006). Measuring the Performance of the Hollow State. Washington, DC, Georgetown University Press.

Frønes, I. (2006). "Theorizing Indicators: On Indicators, Signs and Trends." Social Indicators Research **83**(1): 5-23.

Game, C. (2006). "Comprehensive Performance Assessment in English Local Government." International Journal of Productivity and Performance Management **55**(6): 466-479.

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

Geedes, M. and S. Martin (2000). "The Policy and Politics of Best Value: Currents, Crosscurrents and Undercurrents in the New Regime." Policy & Politics **28**(3): 379–395.

Geys, B. and W. Moesen (2009). "Measuring Local Government Technical (In)Efficiency An Application and Comparison of FDH, DEA, and Econometric Approaches." Public Performance & Management Review **32**(4): 499-513.

Gordon, T., M. Fischer, et al. (2002). "A Comparative Empirical Examination of Extent of Disclosure by Private and Public Colleges and Universities in the United States". Journal of Accounting and Public Policy **21**(3): 235-275.

Grosskopf, S., K. J. Hayes, et al. (2001). "On the Determinants of School District Efficiency: Competition and Monitoring." Journal of Urban Economics **49**(3): 453-478.

Guindon, M. and F. Bellavance (2004). Rapport concernant la catégorisation des organismes municipaux et l'utilisation des facteurs d'influence (Phase III). Montréal, QC, Corporation des officiers municipaux agréés du Québec**:** 26p.

Gulick, L. (2004). Note on the Theory of Organization. Classics of Public Administration. Fifth edition. Shafritz, J.M., Hyde, A.C. and Parkes, S.J. *eds.*, Belmont, CA, Wadsworth/Thomson: 90-98.

Gutiérrez-Romero, R., D. Haubrich, et al. (2008). "The Limits of Performance Assessments of Public Bodies: External Constraints in English Local Government." Environment and Planning C: Government and Policy **26**(4): 767-787.

Guven-Uslu, P. and L. Conrad (2008). "Uses of Management Accounting Information for Benchmarking in NHS Trusts." Public Money & Management **28**(4): 239-246.

Halachmi, A. (2002). "A Brief Note on the Methodology of Measuring Productivity of Services at the Local Level." International Journal Organization Theory & Behavior **5**(3/4): 215–218.

Halachmi, A. (2002). "Performance Measurement: A Look at Some Possible Dysfunctions." Work Study **51**(5): 230-239.

Halachmi, A. (2005). "Performance Measurement: Test The Water Before You Dive In." International Review of Administrative Sciences **71**(2): 255-266.

Hall, J. L. (2007). "Implications of Success and Persistence for Public Sector Performance." Public Organization Review **7**(3): 281-297.

Hamilton, E. K. (1972). "Productivity: The New York City Approach." Public Administration Review **32**(6): 784-795.

Harbour, J. L. and J. L. Marble (2005). "How Performance Improves." Performance Improvement **44**(8): 14-19.

Hatry, H. (1976). "Approaches to Productivity Measurement and Program Evaluation." Public Productivity Review **1**(3): 21-28.

Hatry, H. P. (1980). "Performance Measurement Principle and Techniques: An Overview for Local Government." Public Productivity Review **4**(4): 312-339.

Hatry, H. P. (2006). Performance measurement: Getting results. Second edition.Washington, D.C.: Urban Institute Press.

Hatry, H. (2008). Epilogue: the Many Faces of Use. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 227-240.

Haubrich, D. and I. McLean (2006). "Assessing Public Service Performance in Local Authorities Through CPA - A Research Note on Deprivation." National Institute Economic Review **197**(1): 93-105.

Hauner, D. and A. Kyobe (2008). Determinants of Government Efficiency. IMF Working Paper. Washington, DC**:** 1-25.

Heinrich, C. J. (1999). "Do Government Bureaucrats Make Effective Use of Performance Management Information?" Journal of Public Administration Research and Theory **9**(3): 363-393.

Hendrick, R. (2004). "Assessing and Measuring the Fiscal Health of Local Governments: Focus on Chicago Suburban Municipalities." Urban Affairs Review **40**(1): 78-114.

Hill, H. and M. Andrews (2005). "Reforming Budget Ritual and Budget Practice: The Case of Performance Management Implementation in Virginia." International Journal of Public Administration **28**(3/4): 255-272.

Higgins, P. (2005). "Performance and User Satisfaction Indicators in British Local Government." Public Management Review **7**(3): 445-464.

Hinton, M., G. Francis, et al. (2000). "Best Practice Benchmarking in the UK." Benchmarking **7** (1): 52-61.

HM Treasury (2003). Setting Key Targets for Executive Agencies: A Guide. H.M. Treasury. Norwich, U.K.**:** 1-43.

Ho, A. T.-K. (2008). Reporting Public Performance Information: the Promise and Challenges of Citizen Involvement. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 192-210.

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

Ho, A. T.-K. and A. Y. Ni (2005). "Have Cities Shifted to Outcome-Oriented Performance Reporting? - A Content Analysis of City Budgets." Public Budgeting & Finance **25**(2): 61-83.

Holloway, J., G. Francis, et al. (1999). "A Vehicle for Change? A Case Study of Performance Improvement in the "New" Public Sector." International Journal of Public Sector Management **12**(4): 351-365.

Holzer, M. and E. Charbonneau (2008). Public Management & Administration Illustrated. Newark, NJ, Rutgers-Newark School of Public Affairs and Administration.

Holzer, M. and K. Yang (2004). "Performance Measurement and Improvement: An Assessment of the State of the Art." International Review of Administrative Sciences **70**(1): 15-31.

Hood, C. (2007). "Public Service Management by Numbers: Why Does it Vary? Where Has it Come From? What Are the Gaps and the Puzzles?" Public Money & Management **27**(2): 95-102.

Hoogenboezem, J. A. and D. B. Hoogenboezem (2005). "Coping with Targets: Performance Measurement in the Netherlands Police." International Journal of Productivity and Performance Management **54**(7): 568-578.

Hougaard, J. L., D. Kronborg, et al. (2004). "Improvement Potential in Danish Elderly Care." Health Care Management Science **7**(3): 225-235.

Jacobs, R. and M. Goddard (2007). "How Do Performance Indicators Add Up? An Examination of Composite Indicators in Public Services." Public Money & Management **27**(2): 103-110.

Jamali, D. (2007). "A Study of Customer Satisfaction in the Context of a Public Private Partnership." International Journal of Quality & Reliability Management **24**(4): 370-385.

James, O. and P. John (2006). "Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments." Journal of Public Administration Research and Theory **17**(4): 567-580.

Jansen, E. P. (2008). "New Public Management: Perspectives on Performance and the Use of Performance Information." Financial Accountability & Management **24**(2): 169-191.

Jas, P. and C. Skelcher (2005). "Performance Decline and Turnaround in Public Organizations: A Theoretical and Empirical Analysis." British Journal of Management **16**(3): 195-210.

Jeong, H.-S. and R. Brower (2008). "Extending the Present Understanding of Organizational Sensemaking: Three Stages and Three Contexts." Administration & Society **40**(3): 223-252.

Johansson, T. and S. Siverbo (2009). "Explaining the Utilization of Relative Performance Evaluation in Local Government: A Multi-Theoritical Study Using Data From Sweden." Financial Accountability & Management **25**(2): 197-224.

Johnsen, Å. (1999). "Implementation Mode and Local Government Performance Measurement: A Norwegian Experience." Financial Accountability & Management **15**(1): 41-66.

Johnsen, Å. (2005). "What Does 25 Years of Experience Tell Us About the State of Performance Measurement in Public Policy and Management?" Public Money & Management **25**(1): 9-17.

Johnsen, Å. (2008). Performance Information in Educational Policy Making. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 157-173.

Johnsen, Å. and J. Vakkuri (2006). "Is There a Nordic Perspective on Public Sector Performance Measurement?" Financial Accountability & Management **22**(3): 291-308.

Jones, D. S. (2001). "Performance Measurement and Budgetary Reform in the Singapore Civil Service." Journal of Public Budgeting, Accounting & Financial Management **13**(4): 485-511.

Jordan, M. M. and M. Hackbart (2005). "The Goals and Implementation Success of State Performance-Based Budgeting." Journal of Public Budgeting, Accounting & Financial Management **17**(4): 471-487.

Jreisat, J. E. (1987). "Productivity Measurement: Trial and Error in St.Petersburg." Public Productivity Review **11**(2): 3-18.

Kampen, J. K. (2007). "The Impact of Survey Methodology and Context on Central Tendency, Nonresponse and Associations of Subjective Indicators of Government Performance." Quality & Quantity **41**(6): 793–813.

Kassel, D. S. (2008). "Performance, Accountability, and the Debate over Rules." Public Administration Review **68**(2): 241-252.

Kaufman, R. (2006). "Failure: What It Is and How to Invite It." Performance Improvement **45**(7): 9-12.

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

Kelly, J. M. and D. Swindell (2002). "Service Quality Variation Across Urban Space: First Steps Toward a Model of Citizen Satisfaction." Journal of Urban Affairs **24**(3): 271-288.

Kelman, S. and J. N. Friedman (2009). "Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service." Journal of Public Administration Research and Theory **19**(4): 917-946.

Kenis, P. (2006). Performance Control and Public Organizations. Public Service Performance: Perspectives on Measurement and Management. G. A. Boyne, K. J. Meier, L. J. J. O'Toole and R. M. Walker. New York, NY, Cambridge University Press**:** 113-129.

Kent, R. (2009). "Rethinking Data Analysis –Part One: The limitations of Frequentist Approaches." International Journal of Market Research **51**(1): 51-69.

Kerr, C. A., J. C. Glass, et al. (1999). "Best-Practice Measures of Resource Utilization for Hospitals: A Useful Complement in Performance Assessment." Public Administration **77**(3): 639-650.

Kinney, A. S. (2008). "Current Approaches to Citizen Involvement in Performance Measurement and Questions they Raise." National Civic Review **97**(1): 46-54.

Koopmans, T.C. (1951). "Efficient Allocation of Resources." Econometrica **19**(4):455-465.

Kouzmin, A., E. Loffler, et al. (1999). "Benchmarking and Performance Measurement in Public Sectors Towards Learning for Agency Effectiveness." International Journal of Public Sector Management **12**(2): 121-144.

Lægreid, P., P. G. Roness, et al. (2008). Performance Information and Performance Steering: Integrated System or Loose Coupling? Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 42-57.

Lawton, A., D. McKevitt, et al. (2000). "Coping with Ambiguity: Reconciling External Legitimacy and Organizational Implementation in Performance Measurement." Public Money & Management **20**(3): 13-19.

Li, K.H., Raghunathan, T. E., and D.B. Rubin (1991). "Large-Sample Significance Levels from Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution." Journal of the American Statistical Association **86**(416): 1065-1073.

Liederbach, J., E. J. Fritsch, et al. (2008). "Exploring the Limits of Collaboration in Community Policing: A Direct Comparison of Police and Citizen Views." International Journal of Police Strategies & Management **31**(2): 271-291.

Lied, T. R. (2001). "Small Hospitals and Performance Measurement: Implications and Strategies." International Journal of Health Care Quality Assurance **14**(4/5): 168-173.

Lindblad, M. R. (2006). "Performance Measurement in Local Economic Development." Urban Affairs Review **41**(5): 646-672.

Locke, E. A. and G. P. Latham (2002). "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey." American Psychologist **57**(9): 705-717.

Lonti, Z. and R. Gregory (2007). "Accountability or Countability? Performance Measurement in the New Zealand Public Service, 1992–2002." Australian Journal of Public Administration **66**(4): 468–484.

Loveday, B. (2006). "Policing Performance: The Impact of Performance Measures and Targets on Police Forces in England and Wales." International Journal of Police Science & Management **8**(4): 282–293.

Lu, Y. (2007). "Performance Budgeting: The Perspective of State Agencies." Public Budgeting & Finance **27**(4): 1-17.

Lu, Y. (2008). "Managing the Design of Performance Measures: The Role of Agencies." Public Performance & Management Review **32**(1): 7-24.

MacPherson, M. (2001). "Performance Measurement in Not-for-Profit and Public-Sector Organisations." Measuring Business Excellence **5**(2): 13-17.

Mante, B. and G. 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.

Marcuccio, M. and I. Steccolini (2009). "Patterns of Voluntary Extended Performance Reporting in Italian Local Authorities." International Journal of Public Sector Management **22**(2): 146-167.

Marsden, G. and P. Bonsall (2006). "Performance Targets in Transport Policy." Transport Policy **13**(3): 191-203.

Matthews, J. W. and A. G. Endress (2008). "Performance Criteria, Compliance Success, and Vegetation Development in Compensatory Mitigation Wetlands." Environmental Management **41**(1): 130-141.

McAdam, R. and L. O'Neill (2002). "Evaluating Best Value through Clustered Benchmarking in UK Local Government: Building Control Services." International Journal of Public Sector Management **15**(6/7): 438-457.

McGowan, R. P. and T. H. Poister (1985). "Impact of Productivity Measurement Systems on Municipal Performance." Policy Studies Review **4**(3): 532-540.

Medori, D. and D. Steeple (2000). "A Framework for Auditing and Enhancing Performance Measurement Systems" International Journal of Operations & Production Management **20**(5):520-533.

Meier, K. J. and J. Bohte (2003). "Not with a Bang, but a Whimper: Explaining Organizational Failures." Administration & Society **35**(1): 104-121.

Meijer, A. J. (2007). "Publishing Public Performance Results on the Internet: Do Stakeholders Use the Internet to Hold Dutch Public Service Organizations to Account?" Government Information Quarterly **24**(1): 165-185.

Mellahi, K. and A. Wilkinson (2004). "Organizational Failure: A Critique of Recent Research and a Proposed Integrative Framework." International Journal of Management Reviews **5/6**(1): 21-41.

Melkers, J. and K. Willoughby (2005). "Models of Performance-Measurement Use in Local Governments: Understanding Budgeting, Communication, and Lasting Effects." Public Administration Review **65**(2): 180-190.

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

Meyer, M. W. and L. G. Zucker (1989). Permanently Failing Organizations. Newbury Park, CA, Sage Publications.

Miller, T. I. and M. A. Miller (1991). "Standards of Excellence: U.S. Residents' Evaluations of Local Government Services." Public Administration Review **51**(6): 503-513.

Miller, D. (1994). "What Happens After Success: The Perils of Excellence." Journal of Management Studies **31**(3): 325-358.

Miller, W. H., B. Kerr, et al. (2008). "School Performance Measurement: Politics and Equity." American Review of Public Administration **38**(1): 100-117.

Ministère des Affaires municipales, du Sport et du Loisir (2004). Les Indicateurs de Gestion Municipaux: Un regard neuf sur notre municipalité, Gouvernement du Québec, Québec.

Ministère des Affaires municipales et Régions (2005). Facteurs d'influence prédéterminés http://www.mamr.gouv.qc.ca/finances/fina_indi_resu_fact.htm
Accessed December 10th, 2007.

Ministère des Affaires municipales et Régions (2008). Décret de population [Population Decree]
http://www.mamr.gouv.qc.ca/finances/fina_indi_tran.asp#tramsmission
Accessed *March 29th, 2008.*

Minnesota Department of Transportation (2003). 2003 Statewide Transportation Plan
Mn/DoT official website
http://www.dot.state.mn.us/planning/stateplan/03plan.html
Accessed October 1st, 2009.

Minnesota Department of Transportation (2009). Minnesota Statewide Transportation Policy Plan: 2009-2028
Mn/DoT official website
http://www.dot.state.mn.us/planning/stateplan/Final%20Plan%20Documents/Policy%20Plan/Entire/Minnesota%20Statewide%20Transportation%20Policy%20Plan_2009-2028.pdf
Accessed October 1st, 2009.

Moore, A., J. Nolan, et al. (2005). "Putting Out the Trash: Measuring Municipal Service Efficiency in U.S. Cities." Urban Affairs Review **41**(2): 237-259.

Moullin, M., J. Soady, et al. (2007). "Using the Public Sector Scorecard in Public Health." International Journal of Health Care Quality Assurance **20**(4): 281-289.

Moynihan, D. P. (2005). "Managing for Results in an Impossible Job: Solution or Symbol?" International Journal of Public Administration **28**(3/4): 213-231.

Moynihan, D. (2008a). Advocacy and Learning: An Interactive-Dialogue approach to Performance Information Use. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 25-41.

Moynihan, D. P. (2008b). The Dynamics of Performance Management: Constructing Information and Reform. Washington, DC, Georgetown University Press.

Moxham, C. (2009). "Performance Measurement: Examining the Applicability of the Existing Body of Knowledge to Nonprofit Organisations." International Journal of Operations & Production Management **29**(7): 740-763.

Myhre, B. A. (2008). "Public Employees as Partners in Performance: Lessons from the Field." National Civic Review **97**(1): 29-34.

New York City *City Performance Reporting* (2009). Agency Performance Report of the Department of Homeless Services
Mayor's Office of Operations official website
http://www.nyc.gov/html/ops/cpr/html/home/home.shtml
Accessed October 6th, 2009.

Niskanen, W. (1971). Bureaucracy and Representative Government. Aldine, IL, Chicago University Press.

O'DonnellL, T. K. and D. L. Galat (2008). "Evaluating Success Criteria and Project Monitoring in River Enhancement Within an Adaptive Management Framework." Environmental Management **41**(1): 90-105.

Ormerod, P. (2005). Why Most Things Fail: Evolution, Extinction & Economics. Hoboken, NJ, Wiley.

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

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

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

Palmer, T. B. and J. C. Short (2001). "Why Do Goliaths Fall? Performance Referents in Successful Organizations." Corporate Reputation Review **4**(3): 210-222.

Perrin, B. (1998). "Effective Use and Misuse of Performance Measurement." American Journal of Evaluation **18**(3): 367-379.

Pina, V. and L. Torres (2001). "Analysis of the Efficiency of Local Government Services Delivery. An Application to Urban Public Transport." Transportation Research **35**(10): 929-944.

Piotrowski, S. J. and D. H. Rosenbloom (2002). "Nonmission-Based Values in Results-Oriented Public Management: The Case of Freedom of Information." Public Administration Review **62**(6): 643-657.

Piotrowski, S. J. (2003). Governmental Transparency and the National Performance Review: Implementing the Freedom of Information Act. School of Public Affairs. Washington, D.C., American University**:** 1-206.

Plant, T. and J. Douglas (2006). "The Performance Management Continuum in Municipal Government Organizations." Performance Improvement **45**(1): 43-48.

Poister, T. H. and G. Streib (1999). "Performance Measurement in Municipal Government: Assessing the State of the Practice." Public Administration Review **59**(4): 325-335.

Poister, T. H. and J. C. Thomas (2007). "The Wisdom of Crowds: Learning from Administrators' Predictions of Citizen Perceptions." Public Administration Review **67**(2): 279-289.

Pollanen, R. M. (2005). "Performance Measurement in Municipalities: Empirical Evidence in Canadian Context." International Journal of Public Sector Management **18**(1): 4-24.

Pongatichat, P. and R. Johnston (2008). "Exploring Strategy-Misaligned Performance Measurement." International Journal of Productivity and Performance Management **57**(3): 207-222.

Pritchard, R. D., S. S. Culbertson, et al. (2009). "Improving Performance in a Swedish Police Traffic Unit: Results of an Intervention." Journal of Criminal Justice **37**(1): 85-97.

Public Administration Select Committee (2003). On Target? Government By Measurement. Fifth Report of Session 2002–03. London, UK, House of Commons**:** 1-53.

Purdue, D. (2005). "Performance Management for Community Empowerment Networks." Public Money & Management **25**(2): 123-130.

Raaum, 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.

Radin, B. A. (1998). "The Government Performance and Results Acts (GPRA): Hydra-Headed Monster or Flexible Management Tool?" Public Administration Review **58**(4): 307-316.

Radnor, Z. (2008). Hitting the Target and Missing the Point? Developing an Understanding of Organizational Gaming. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 94-105.

Radnor, Z. (2008b). "Muddled, Massaging, Manoeuvring or Manipulated?: A Typology of Organisational Gaming." International Journal of Productivity and Performance Management **57**(4): 316-328.

Rantanen, H., H. I. Kulmala, et al. (2007). "Performance Measurement Systems in the Finnish Public Sector." International Journal of Public Sector Management **20**(5): 415-433.

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

Richards, C. E. and R. Height (1988). "Accountability and Performance Incentives for School Principals." Public Productivity & Review **12**(1): 15-27.

Ridley, C. E. (1927). Measuring Municipal Government: Suggested Standards for Measuring the Results of Fire, Health, Police and Public Works Departments. Syracuse, NY, School of Citizenship and Public Affairs, Syracuse University.

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

Rivenbark, W. C., D. N. Ammons, et al. (2007). "Benefiting from Comparative Performance Statistics in Local Government." Popular Government **72**(3): 34-42.

Roberts, G. E. (1996). "A Case Study in Performance Appraisal System Development: Lessons from a Municipal Police Department." American Review of Public Administration **26**(3): 361-385.

Robson, I. (2005). "Implementing a Performance Measurement System Capable of Creating a Culture of High Performance." International Journal of Productivity and Performance Management **54**(2): 137-145.

Roche, S. (2008). "Performance Management in France: A Police or an Electoral Issue?" Policing: A Journal of Policy and Practice **2**(3): 331-339.

Rogers, M. K. (2006). Explaining Performance Measurement Utilization and Benefits: An Examination of Performance Measurement Practices in Local Governments. Department of Public Administration. Raleigh, NC, North Carolina State University**:** 1-284.

Rogerson, P. (1995). "Performance Measurement and Policing: Police Service or Law Enforcement Agency?" Public Money & Management **15**(4): 25-30.

Rosenbloom, D. H. and S. J. Piotrowski (2005). "Outsourcing the Constitution and Administrative Law Norms." American Review of Public Administration **35**(2): 103-121.

Rosentröm, U. and S. Kyllönen (2007). "Impacts of a Participatory Approach to Developing National Level Sustainable Development Indicators in Finland." Journal of Environmental Management **84**(3): 282–298.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. J. Wiley & Sons, New York, NY.

Rubin, D.B. (1996). "Multiple Imputation after 18+ Years." Journal of the American Statistical Association, 91(434): 473-489.

Rubin, E. (2005). "The Myth of Accountability and the Anti-Administrative Impulse." Michigan Law Review **103**(8):2073-2136.

Salminen, M. and M.-L. Viitala (2006). Handbook on Performance Management. Public Management Department. Helsinki, FI, Ministry of Finance**:** 1-146.

Sarrico, C. S. and M. J. Rosa (2009). "Measuring and Comparing the Performance of Portuguese Secondary Schools: A Confrontation between Metric and Practice Benchmarking." International Journal of Productivity and Performance Management **58**(8): 767-786.

Savoie, D.J. (1999). Governing from the Centre: The Concentration of Power in Canadian Politics. Toronto, ON, University of Toronto Press.

Schiffel, L. and K. Smith (2006). "How Does the Media Report on Government Performance?" Journal of Government Financial Management **55**(3): 18-23.

Schneider, E. C. and A. M. Epstein (1998). "Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery." Journal of the American Medical Association **279**(20): 1638-1642.

Schweigert, F. J. (2006). "The Meaning of Effectiveness in Assessing Community Initiatives." American Journal of Evaluation **27**(4): 416-436.

Scollay, S. J. and S. T. Everson (1985). "Measuring School Improvement: A Few Experientially Based Words of Caution." Urban Review **17**(3): 201-209.

Scorsone, E. A. (2008). "New Development: What are the Challenges in Transferring Lean Thinking to Government?" Public Money & Management **28**(1): 61-64.

Simon, H. A. and C. E. Ridley (1938). "Trends in Municipal Reporting." Public Opinion Quarterly **2**(3): 465-468.

Simon, H. A. (1997). Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations. Fourth edition. New York, NY, Free Press.

Siverbo, S. and T. Johansson (2006). "Relative Performance Evaluation in Swedish Local Government." Financial Accountability & Management **22**(3): 271-290.

Smith, P. (1995). "On the Unintended Consequences of Publishing Performance Data in the Public Sector." International Journal of Public Administration **18**(2-3): 277-310.

Smith, P. C. (2005). "Performance Measurement in Health Care: History, Challenges and Prospects." Public Money & Management **25**(4): 213-220.

Smyth, R. L., M. C. Watzin, et al. (2007). "Defining Acceptable Levels for Ecological Indicators: An Approach for Considering Social Values." Environmental Management **39**(3): 301-315.

Sohl, S., M. T. Peddle, et al. (2009). "Measuring the Financial Position of Municipalities: Numbers Do Not Speak for Themselves." Public Budgeting & Finance **29**(3): 74-96.

Spendolini, M.J. (1992). The Benchmarking Book, American Management Association, New York, NY.

Spottiswoode, C. (2000). Improving Police Performance: A New Approach to Measuring Police Efficiency. Public Services Productivity Panel. London, U.K., HM Treasury**:** 1-48.

Stein, B. (2007). "Budgeting for Outcomes: The Right Results for the Right Price." Government Finance Review **23**(5): 58-60.

Steinhardt, B. (2008). Lessons Learned for the Next Administration on Using Performance Information to Improve Results. Washington, DC, United States Government Accountability Office**:** 1-28.

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

Stone, M. and J. Davis (2007). "Weighting and Scoring' in Theory and in Practice." Public Money & Management **27**(3): 215-222.

Straight, R. L. (2000). "Performance Metrics: Avoiding the Pitfalls." Public Administration Quarterly **23**(4): 495-516.

Streib, G. and T. H. Poister (1999). "Assessing the Validity, Legitimacy, and Functionality of Performance Measurement Systems in Municipal Governments." American Review of Public Administration **29**(2): 107-123.

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

Talley, W. K. (1986). "A Comparison of Two Methodologies for Selecting Transit Performance Indicators." Transportation **13**(3): 201-210.

Taylor, J. (2006). "Performance Measurement in Australian and Hong Kong Government Departments." Public Performance & Management Review **29**(3): 334–357.

Taylor, J. (2009). "Strengthening the Link between Performance Measurement and Decision Making." Public Administration **87**(4): 853-871.

Thanassoulis, E. (1995). "Assessing Police Forces in England and Wales Using Data Envelopment Analysis." European Journal of Operational Research **87**(3): 641-657.

Tilbury, C. (2006). "Accountability via Performance Measurement: The Case of Child Protection Services." Australian Journal of Public Administration **65**(3): 48-61.

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

Trosa, S. and S. Williams (1996). Benchmarking in Public Sector Performance Management, Performance Measurement in Government. Occasional Paper No. 9, OECD, Paris, FR.

Try, D. and Z. Radnor (2007). "Developing an Understanding of Results-Based Management through Public Value Theory." International Journal of Public Sector Management **20**(7): 655-673.

Usher, C. L. and G. C. Cornia (1981). "Goal Setting and Performance Assessment in Municipal Budgeting." Public Administration Review **41**(2): 229-235.

Van Dooren, W. (2008). Nothing New Under the Sun? Change and Continuity in the Twentieth-Century Performance Movements. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 11-23.

Van de Walle, S. (2009). "International Comparisons of Public Sector Performance: How to Move Ahead?" Public Management Review **11**(1): 39-56.

Van De Walle, S. and G. Bouckaert (2007). "Perceptions of Productivity and Performance in Europe and The United States." International Journal of Public Administration **30**(11): 1123–1140.

Van De Walle, S. and W. Van Dooren (2008). Introduction: Using Public Performance Information. Performance Information in the Public Sector: How It Is Used. W. Van Dooren and S. Van De Walle. Houndmills, UK, Palgrave Macmillan**:** 1-8.

Van Peursem, K. A., M. J. Pratt, et al. (1995). "Health Management Performance: A Review of Measures and Indicators." Accounting, Auditing & Accountability Journal **8**(5): 34-70.

Van Ryzin, G.G. and E. Charbonneau (forthcoming). "Public Service Use and Perceived Performance: An Empirical Note on the Nature of the Relationship." Public Administration

Voyer, P. (1999). Tableaux de bord de gestion et indicateurs de performance. 2e édition, Québec, QC, Presses de l'Université du Québec.

Wadman, R. C. and J. C. DeLadurantey (1984). "Overcoming Limitations to Police Productivity Measurement: The Omaha Experience." Public Productivity Review **8**(3): 225-235.

Wagner, J. A. I. and R. Z. Gooding (1997). "Equivocal Information and Attribution: An Investigation of Patterns of Managerial Sensemaking " Strategic Management Journal **18**(4): 275-286.

Waldo, D. (1984). The Administrative State. Second edition. New York, NY, Holmes and Meier Publishers.

Wall, A. and G. Martin (2003). "The Disclosure of Key Performance Indicators in the Public Sector: How Irish Organizations are Performing." Public Management Review **5**(4): 491-509.

Wallace, R. J. (1977). "Productivity Measurement in the Fire Service." Public Productivity Review **2**(3): 12-36.

Wang, X. (2000). "Performance Measurement in Budgeting: A Study of County Governments." Public Budgeting & Finance **20**(3): 102-118.

Wang, X. and E. Berman (2001). "Hypotheses about Performance Measurement in Counties: Findings from a Survey." Journal of Public Administration Research and Theory **11**(3): 403-427.

Wang, X. (2002). "Perception and Reality in Developing and Outcome Performance Measurement System." International Journal of Public Administration **25**(6): 805-829.

Wedding, G. C. and D. Crawford-Brown (2007). "Measuring Site-Level Success in Brownfield Redevelopments: A Focus on Sustainability and Green Building." Journal of Environmental Management **85**(2): 483–495.

Weick, K. E. (1995). Sensemaking in Organizations. Thousand Oaks, CA, Sage Publications.

Wholey, J. S. and H. P. Hatry (1992). "The Case for Performance Monitoring." Public Administration Review **52**(6): 604-610.

Willcox, W. F. (1896). "Methods of Determining the Economic Productivity of Municipal Enterprises " American Journal of Sociology **2**(3): 378-391.

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**:** 1-246.

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

Wilson, J. Q. (1993). The Problem of Defining Agency Success. Performance Measures for the Criminal Justice System: Discussion Papers from the BJS-Princeton Project. J. Q. Wilson. Washington DC, DIANE Publishing Company**:** 156-164.

Wilson, D. (2004). "Which Ranking? The Impact of a 'Value-Added' Measure of Secondary School Performance." Public Money & Management **24**(1): 37-45.

Wisniewski, M. and D. Stewart (2004). "Performance Measurement for Stakeholders: The Case of Scottish Local Authorities." International Journal of Public Sector Management **17**(2): 222-233.

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

Yang, K. and M. Holzer (2006). "The Performance-Trust Link: Implications for Performance Measurement." Public Administration Review **66**(1): 114-126.

Yang, K., J. Y. Hsieh, et al. (2009). "Contracting Capacity and Perceived Contracting Performance: Nonlinear Effects and the Role of Time." Public Administration Review **69**(4): 681-696.

Yockey, M. D. and S. M. Kruml (2009). "Everything is Relative, but Relative to What? Defining and Identifying Reference Points." Journal of Business and Management **15**(1): 95-109.

Young, M. (1991). An Inside Job: Policing and Police Culture.Oxford, UK, Clarendon Press.

Zafra-Gómez, J. L., A. M. López-Hernández, et al. (2009). "Developing a Model to Measure Financial Condition in Local Government: Evaluating Service Quality and Minimizing the Effects of the Socioeconomic Environment: An Application to Spanish Municipalities." American Review of Public Administration **39**(4): 425-449.

Zedlewski, E. W. (1979). "Performance Measurement in Public Agencies: The Law Enforcement Evolution." Public Administration Review **39**(5): 488-493.

APPENDIX

Identification of the Municipality

Municipality's name:
You municipal geographical identification number: (*if you do not know your municipality's GIN, you can find it on MAMROT's website : http://www.mamrot.gouv.qc.ca/ , see : « Répertoire des municipalités »*.)

1. Data collection and reporting municipal management indicators has been mandatory for all municipalities since 2003. According to your observations, what is the utilization level in your municipality?

In my municipality, the mandatory management indicators… *(Check one box that matches your perception.)*

☐ … are used very often
☐ … are used often
☐ … are seldom used
☐ … are not used. If you answered "are not used", briefly indicate why below and move directly to question 5 for the remainder of the survey.

2. **From what you have observed, indicate what are the reasons for which management indicators are used in your municipality:** *(check all boxes that apply.)*

☐ To prepare budgets, including resources allocations or discussion of resources reallocations.
☐ In establishing contracts for services (ex. snow removal).
☐ Managing operations or routine decisions (e.g., scheduling activities)
☐ Evaluation to establish underlying reasons for results
☐ Specific performance improvement initiatives (e.g., investments, technical assistance, training, operations improvements)
☐ To provide feedback to managers and employees.
☐ To report to elected officials.
☐ To report to citizens, citizen groups or to inform the medias (ex.: the mayoral report on financial health).
☐ Others (elaborate):

Indicate if since their implementation, mandatory management indicators for the different functions and activities have been explicitly mentioned [appeared] in the preparation of the budget and the annual report on the financial situation in your municipality. *(check all boxes that apply.)*

Function & Activity	**Annual budget**		**Annual report on the financial situation**	
	Yes	**No**	**Yes**	**No**
Roads Municipal roadway system	☐	☐	☐	☐
Roads Snow removal	☐	☐	☐	☐
Public Hygiene Water supply, treatment and distribution	☐	☐	☐	☐
Public Hygiene Used water treatment and sewage systems	☐	☐	☐	☐
Global Financial Health	☐	☐	☐	☐
Human Resources	☐	☐	☐	☐

3. For each mandatory indicators presented in the table below, first indicate if you ever compared the value of the indicator, and second, if you ever established targets for this indicator. If you checked "yes" for either one of these questions, elaborate by checking all boxes that applies in the table.

Function & Activity	Indicator	Have you ever compared the results?		If your answer is *yes*, which elements were used while comparing your results?				Have you ever established targets?		If your answer is *yes*, which elements were used while establishing your targets?			
		Yes	No	Previous results	Quartiles from CPEGM* annual reports	Results from comparable municipalities	Others	Yes	No	Previous results	Quartiles from CPEGM* annual reports	Results from comparable municipalities	Others
Roads Municipal roadway system	Cost of the municipal roadway system, per km	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Roads Snow removal	Cost of snow removal, per km	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Public Hygiene Water supply, treatment and distribution	Number of breaks, per 100 km of pipes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Cost of distribution, per km of pipes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Cost of supply and treatment of water per m^3	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Cost of water distribution per m^3	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Public Hygiene Used water treatment and sewage systems	Cost of treatment of used water per m^3	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Cost of sewage system per km of pipes	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Global Financial Health	Percentage of debt service	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Indebtedness percentage	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Human Resources	Training effort per employee	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Percentage of training cost, compared to total payroll	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Average length of health-related leaves of absence	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Potential retirement rate	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

* CPEGM : *Centre de promotion de l'excellence en gestion municipale* [Centre for the Promotion of Excellence in Municipal Management] (http://neumann.hec.ca/cpegm/)

4.In your municipality, do managers use 'tableaux de bord' [something close to dashboards, only present in France and Quebec] (computerized or manual) or any other developed system specifically designed for the management of services under your responsibility?

☐ Yes. Precise your answer by checking all boxes that apply:

- ☐ Manual 'tableaux de bord'.
- ☐ 'Tableaux de bord' on Excel.
- ☐ Computerized 'Tableaux de bord' other than Excel.
- ☐ Other manual system to display indicators.
- ☐ Other computerized system to display indicators on Excel.
- ☐ Other computerized system to display indicators.

☐ No.

5.The following table presents a series of scenarios related to hypothetical results on a management indicator. Indicator values have to be interpreted in their context by managers: indicators do not speak by themselves. For each of the following scenarios described below, indicate how you would interpret the value of that indicator. There are no right or wrong answers. We are seeking your perceptions.

While analyzing the results of management indicators in your municipality, how would you interpret the following statements? *(Check only one box for each scenario. Darkened boxes indicate unavailable selections.)*

Scenarios	**A success**	**Satisfactory**	**Un-satisfactory**	**A failure**
A result *worse* than last year performance?		☐	☐	☐
A result *worse* than a predetermined target?		☐	☐	☐
A result *worst* than other municipalities of comparable size.		☐	☐	☐
A result *worse* than other municipalities with comparable characteristics.		☐	☐	☐
The *same* result as last year performance?	☐	☐	☐	☐
A result *meeting* a predetermined target?	☐	☐		
A result considered *average* to other municipalities of comparable size.	☐	☐	☐	
A result considered *average* to other municipalities with comparable characteristics.	☐	☐	☐	
A result *better* than last year performance?	☐	☐		
A result *better* than a predetermined target?	☐	☐		
A result *better* than other municipalities of comparable size.	☐	☐		
A result *better* than other municipalities with comparable characteristics.	☐	☐		

6. **Some managers identified barriers which would limit the use of management indicators in decision making. Indicate your level of agreement regarding the following statements on the management indicators.** *(Check only one box per statement.)*

Statements	**Agree**	**Somewhat agree**	**Somewhat disagree**	**Disagree**
Management indicators are not considered useful.	☐	☐	☐	☐
Management indicators are not trustworthy.	☐	☐	☐	☐
Management indicators are felt to convey an incomplete picture of the organization.	☐	☐	☐	☐
We fear that management indicators are misunderstood and misinterpreted.	☐	☐	☐	☐
We do not know how to integrate management indicators into decision making.	☐	☐	☐	☐
We are not able to access data that would enable us to compare our results to similar municipalities.	☐	☐	☐	☐
We lack the time to use management indicators.	☐	☐	☐	☐
We lack the staff with the expertise to work with management indicators.	☐	☐	☐	☐
We lack the computerized tools to gather the detailed data on the management indicators.	☐	☐	☐	☐
We need additional information to use the management indicators.	☐	☐	☐	☐
Our elected officials are uninterested in the management indicators.	☐	☐	☐	☐
Management indicators are seen as a threat.	☐	☐	☐	☐
Management indicators will expose our weaknesses.	☐	☐	☐	☐

7. **Indicate your level of agreement regarding the following statements on the management indicators*, on the current situation in your municipality.*** *(Check only one box per statement*

Statement on the current situation in your municipality	Agree	Somewhat agree	Somewhat disagree	Disagree
There are clear links between the objectives and priorities of our service and those for the municipality as a whole.	☐	☐	☐	☐
Co-ordination and joint working among the different municipal services is a major part of our approach to the organization of services.	☐	☐	☐	☐
The general manager and most managers place the needs of users first and foremost when planning and delivering services.	☐	☐	☐	☐
Strategic planning is generally made in consultation with our external stakeholders (citizen groups, chamber of commerce, etc.).	☐	☐	☐	☐
The elaboration of strategic planning is augmented by the utilization of management indicators.	☐	☐	☐	☐
Working more closely with our citizens is a major part of our approach to service delivery.	☐	☐	☐	☐
Citizens' demands are important in driving service improvement.	☐	☐	☐	☐
Political leadership is important in driving performance improvement.	☐	☐	☐	☐
The general manager is important in guiding decision making to drive performance improvement.	☐	☐	☐	☐
The level of support from the general manager is important in determining the municipality's approach to the use of management indicators.	☐	☐	☐	☐
The level of participation from managers is important in determining the municipality's approach to the use of management indicators.	☐	☐	☐	☐
The municipality's objectives are clearly and widely communicated by managers of different services.	☐	☐	☐	☐

Statement on the current situation in your municipality	Agree	Somewhat agree	Somewhat disagree	Disagree
The annual report on the management indicators is discussed with politicians.	☐	☐	☐	☐
The annual report on the management indicators is discussed with managers of different services.	☐	☐	☐	☐

8. Do you use the management indicators calculation software that can be downloaded freely from MAMROT's website?

☐ Yes. ☐ No.

Information on the respondent

Current function:

General Manager ☐
Treasurer ☐
Other (specify)

Number of years performing the current function: _____
Number of years performing a similar of related function: _____

If you have comments to formulate regarding the management indicators or the present survey, please add them in the space below:

Please save this file if you filled it out electronically.

We recommend that you keep a copy for your archives, in addition to the copy that you will send us.

Please send us the completed survey, in preferably by e-mail to cpegm@hec.ca , or by regular mail at this address:

CPEGM
Bureau 5.447
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine,
Montréal (Québec) H3T 2A7

Thank you very much for your collaboration!

Curriculum Vitae
Étienne Charbonneau

1980	Born, November 16 in Montreal, Quebec, Canada
1998	Graduated from Polyvalente Thérèse-Martin, Joliette, Quebec
1998-2000	Attended the Cégep Régional de Lanaudière à Joliette, Joliette, Quebec
2000	D.E.C. in Human Sciences, Cégep Régional de Lanaudière à Joliette
2001-2006	Attended Université Laval, Quebec City, Quebec
2003	B.A. Political Science, Université Laval
2004-2005	Research Assistant, Center of Analysis of Public Policies, Université Laval
2006	M.A. Policy Analysis, Université Laval
2006-2010	Attended the School of Public Affairs and Administration, Rutgers University, Campus at Newark, New Jersey
2006-2009	Senior Research Associate, National Center for Public Performance, Rutgers University
2008	Article: Charbonneau, E. and N.M. Riccucci (2008). "Beyond the Usual Suspects: An Analysis of the Performance Measurement Literature on Social Equity Indicators in Policing," *Public Performance & Management Review* 31(4):602-619.
2008	Teaching guide: Holzer, M. and E. Charbonneau, *eds.* (2008). "Public Management & Administration Illustrated," published in partnership with the *Division for Public Administration and Development Management* of the *United Nations Department of Economic and Social Affairs* and the *American Society for Public Administration*, 137p. ISBN: 0-942942-07-8
2008-2009	Assistant Director for Resource Development, Public Performance Measurement & Reporting Network, Rutgers University
2009	Instructor in Public Administration, School of Public Affairs and Administration, Rutgers University
2009	Book chapter: Charbonneau, É. (2009). "Talking Like a Tax Collector or a Social Guardian? The Use of Administrative Discourse by U.S. State Lottery Agencies," in L.M. Imbeau *ed., Do They Walk Like They Talk? Speech and Action in Policy Processes*, New York, NY, Springer: 223-240.
2009	Article: Holzer, M., Charbonneau, E. and Y. Kim (2009). "Mapping the Terrain of Public Service Quality Improvement: Twenty-five Years of Trends and Practices in the United States," *International Review of Administrative Sciences* 75(3):403-418.
2009	Article: Charbonneau, E., Riccucci, N.M., Van Ryzin, G.G. and M. Holzer (2009). "The Self-Reported Use of Social Equity Indicators in Urban Police Departments in the United States and Canada," *State and Local Government Review* 41(2): 95-107.
2009-2010	Assistant Editor, Public Performance & Management Review
2010	Ph.D. Public Administration, School of Public Affairs and Administration, Rutgers University, Campus at Newark