University of Strathclyde

Department of Management

Business School

Glasgow, UK

The Utilization of Performance Measurement Systems

Designed to Facilitate Reform in the Education Sector of a

Developing Country

By

Hanine Salem

A Thesis presented in fulfillment of the requirements for the PhD

Degree

2010

COPYRIGHT STATEMENT

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyrights Acts as qualified by the University of Strathclyde Regulation 3.50. Due acknowledgment must always be made of the use of any material contained in, or derived from, this thesis.

Signed: Date: 10.10.2010

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 4   85

Data Collection and Analysis   85

CHAPTER 5   136

Findings and Discussions   136

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Around the world, evidence abounds about interest in organizational performance measurement. As a central element of the New Public Management doctrine, performance measurement has increasingly turned into a concept that is intertwined with initiatives to reform the public sector around the world. Performance measurement is regarded by many as an objective tool capable of fulfilling the goal of accountability of the public sector to its citizens, improving prudent decision making by public practitioners and policy makers, and addressing the needs of professional and modern public sector management.

This research study embarks from identified challenges with implementing and utilizing performance measurement systems in the public sector. In particular, the study embarks from challenges that are related to the utilization of data and information that is produced by these measurement systems. The overall goal of this research is to establish a better understanding of the utilization process and gain valuable insights into the factors that influence this utilization. This study falls into applied research and involves the adoption of case study method.

The following is a list of the most important theoretical and practical contributions made by this research:

(1) The identification and description of fifteen tangible and nontangible factors that influence the utilization of performance measurement data.

(2) Improved understanding of the dynamics of performance measurement systems in countries with embryonic measurement experiences.

(3) Practical implications that can (a) facilitate the jobs of public officials who deal with measurement systems, and, (b) provide consultants active in the field of public

sector reform with valid guidance on the introduction of performance measurement systems.

This thesis will be of interest to policy makers and public officials interested in the adoption of performance measurement systems, public practitioners who seek to benefit from their performance measurement systems, and consultants involved in the design and implementation of reform initiatives that entail performance measurement schemes. The thesis will also be of interest to researchers and scholars in the field of performance measurement systems.

CHAPTER I

INTRODUCTION

1.1   Setting the Research Agenda: Background and Origins of the Study

"In setting the management research agenda it is crucial that problems are posed, addressed and disseminated with a sympathy to the managerial and broader societal context" (Tranfield and Starkey, 1998, p.342).

The phenomenal upsurge of management interest and practices in the field of organizational performance measurement around the world over the past three decades was accompanied by a significant rise of academic interest in the field, as this study will clearly demonstrate. Performance measurement research is considered an emerging academic sub-discipline of management research (Thorpe and Beasley, 2004). The nature and characteristics of management research, and the question about whether management can be called an applied discipline, has been a subject of debate for many years (The Bain Report, 1994). In contrast, performance measurement research is regarded as applied in nature because most of its investigations are characterized by the application of particular performance management tools and the use of the case study method (Thorpe and Beasley, 2004).

This research study is interested in organizational performance measurement systems and the way they operate in public sector reform settings. The impetus for conducting this research study originally stemmed from the researcher's professional experience in public sector reform over the past ten years and her involvement in projects that introduced organizational performance management as a method to institute transparency and accountability to the public. The premise of performance

measurement systems, and the ways these systems can push forward many urgent items on governments' public sector reform agenda, became of growing interest to the researcher.

However, it soon became clear to the researcher that almost all policy makers and public practitioners that she interacted with, even those who were most enthusiastic about measurement systems, were struggling with questions related to the design and use of these systems. Public practitioners were particularly battling with structural and procedural limitations and finding it difficult to get their own staff to buy into the sprouting systems. The researcher noted that most of the answers that governments received from their advisors and consultants at that point seemed to arrive out of pre-packaged solutions or celebrated new measurement frameworks. It was also easy to discern the Balanced Scorecard, which at the time represented the first attempt to address the shortcomings of traditional and single-focused performance measurement systems, as the most recommended approach. It is interesting to note that the researcher's concern about a widespread mechanical, and almost unthinking, tendency to adopt the BSC is shared by other researchers in the field until recent years. Warnings about the need to investigate the BSC as a strategy for performance management in developing countries (Gomes and Liddle, 2009) reflected the continuing concern among researchers about the compatibility and feasibility of this framework in different settings and contexts.

It was also around the same time when the performance measurement literature community started bringing to light challenges associated with the adoption of performance measurement frameworks and the misuse of resulting data and information (for example: de Lancer Julnes, 2001; de Lancer Julnes and Holzer, 2001; Kennerley and Neely, 2002; Van Thiel and Leeuw, 2002).

The researcher became increasingly interested in carrying out a study to help public practitioners design, adopt, and implement better performance measurement systems and, as a result, help governments push forward their reform and modernization agendas. In order to conduct a meaningful and useful research study in the field of performance measurement, the research would need to be epitomized by results and

conclusions that answer to the question of application for practitioners and management. It was also important to the researcher to examine the performance measurement phenomenon in a real life environment and investigate the perspectives of public officials and managers on a number of pertinent issues. Firsthand understanding of the rich experiences of practitioners was thought to offer valuable insights into the implementation dilemma. This would also provide governments and policy makers interested in modernizing their public systems with pertinent recommendations that can potentially carry important policy implications.

The process of identifying a clear and exact focus for this research study evolved over time and certainly continued to do so even after the researcher started work on this study. Although there were several interesting topics surrounding the implementation of performance measurement it was critical to sharpen the focus and scope of the study as soon as possible. Initially, there was an interest in examining the relationship between the implementation of measurement systems and the development of an organization's strategy. There was also an initial interest in the development of a new framework that can guide the design of better performance measurement systems. As the thinking about the most urgent issues evolved, the researcher decided to focus on achieving a better understanding of the implementation process versus venturing into design questions. The rationale was that improved understandings of implementation could later on lead to improved designs of measurement systems. The following chapter will describe in more details the exact focus of this study and how this focus emerged out of a thorough review of the literature.

1.2 The Topic and its Significance

Perhaps few fields are as polarized as the field of performance measurement in the public sector. The discussion in the next chapter sheds light on an evident split in the literature between devoted believers and unwavering adversaries of the notion of measurement in public sector settings. Opinions about performance measurement

range from considering it the promised solution for the ailments of public sector management to testifying to the impossibility of measuring the outputs and outcomes of government efforts.

This intellectual split, however, does not seem to have forestalled the expansion of performance measurement ideas and methods into public organizations around the world. The past few decades have been characterized by a remarkable interest in the topic of performance measurement to the extent that some researchers even promoted the idea that the world lives in the middle of a 'performance revolution' (Neely, 1999). Reports and articles on performance measurement had been reportedly appearing at a rate of 1 every 5 hours of every working day between 1994 and 1998, and by 1996 a new performance measurement book was being published in the United States every two weeks (Neely, 1998). In 2003, Wisniewski and Shafti found over 355,000 references to performance measurement on the Web.

Advocates for performance measurement go as far as making the case for an analogy between the functions that performance measures play in political competitions and prices in market competition. Not only can performance measurement information be used strategically in politics, it is argued, the selective utilization of this information may also play an effective role in the "*creative destruction of the present or future political or managerial status quo*" (Johnsen, 2005: 14-15). The next chapter will discuss selective utilization modes that currently exist in performance measurement markets.

The increasing embracement of performance measurement methods in the public sector is attributed to the continuous interest in the policy ideas and set of arguments that the New Public Management doctrine (NPM) calls for (Van Thiel and Leeuw, 2002; Moynihan, 2006). Perhaps the most critical gist of the NPM promise rests in its promise to utilize performance measurement data to monitor the progress of policy implementation and hold public practitioners accountable for the outcomes of these implementations (Gianakis, 2002). This new paradigm allows public practitioners to more effectively utilize their professional expertise and good

judgment to achieve the aspired outcomes. Performance measurement is hence viewed as a fundamental tool to resolve the typical predicament of effectiveness versus control that faces reform in public sector management (Moynihan, 2006): "*the simultaneous desire for professional and efficient administration and democratically accountable government* "(Gianakis, 2002).

Increasing interest in performance measurement is also attributed to a variety of forces including internal changes in the nature and role of public organizations, external changes related to new demands and rising competition, and improvement in data collection, analysis and review through the power of information technologies (Neely, 1999).

1.2.1   Citation analysis

The performance measurement field is apparently dynamic with its concepts on the radar screen of academic scholars and public practitioners alike (Marr and Schiuma, 2002). However, and despite evidence in the literature in favor of a field that is seemingly healthy and growing, it was prudent that this thesis starts out with an accurate understanding of the state of performance measurement research. It was also important to arrive at an empirically grounded understanding of who the intellectual players are within in the field. Identifying the leading authors, as well as the primary theoretical concepts, at the outset of this research was important to achieve a good level of understanding about the field of performance measurement.

In order to achieve this, a citation analysis of relevant research was conducted for publications between 1980 and November of 2009 with the help of a librarian colleague of the researcher. Citation analysis is used for a variety of reasons including examining the growth of a field or concept as a legitimate focus of academic research by quantifiably evaluating journal titles and publications in particular disciplines. Of particular relevance to this research, a citation analysis is considered a useful method to guide the identification of the theories, papers, and

authors that have the most impact on a particular field (Todorov and Glanzel, 1988). The underlying assumption is that more frequently cited publications, for example, will have a greater influence on a field than publication that are less cited.

At least two prior research works used citation analysis to improve understanding of the performance measurement field. Marr and Schiuma (2002) examined papers from the first and second international conferences on performance measurement. Neely (2005) examined publications from the ISI Web of Science database with the term 'performance measurement' from 1981 till 2005. Although these two publications provide useful information on the status of performance measurement research, the citation analysis in this study used different criteria and datasets to understand the status of the field.

Citation analysis is typically aided by citation indexes, such as ISI's Web of Knowledge Indexes for Science, Social Sciences, and Arts and Humanities. These indexes allow researchers to examine known articles and determine what they cite, who cites them as well as explore the inter-relationships among authors and emergence and development of new fields of inquiry. The dataset used for the analysis in this research was a result of a search across all ISI Web of Science's Citation Indexes (Social Sciences, Sciences, Arts/Humanities) to ensure that any outlying but relevant results were not missed. A decision was taken to exclude the Proceedings databases because of concerns that these would clutter up the results.

In selecting the search terms for the citation analysis, the term "performance measurement" was initially used. However, it soon became clear that there were few problems with doing a simple search of this term. In addition to getting false hits, such as articles about "performance measurements of mutual funds"), the use of this term will require intensive and lengthy review of the results to get rid of false hits. In addition, such a simple search will leave out relevant articles that reflect the *concept* of "performance measurement" but don't actually use the *term* "performance measurement." An important example of this is that such a search will NOT retrieve the following paper by Andy Neely: Designing performance measures: a structured

approach Author(s): Neely A., Richards H., Mills J. *et al*. Source: INTERNATIONAL JOURNAL OF OPERATIONS & PRODUCTION MANAGEMENT Volume: 17   Issue: 11-12   Pages: 1131-+ Published: 1997 Times Cited: 44. The term "performance measurement" does not appear in the title or the abstract of this article, thus is not retrieved by a simple search of "performance measurement."

Consequently, a more complex search was used to capture relevant results. The Web of Knowledge subject categories were used to narrow the results to the concepts of performance measurement as used in this research field. These subject categories were: OPERATIONS RESEARCH & MANAGEMENT SCIENCE; MANAGEMENT; BUSINESS. In addition, more inclusive terms were used such as "performance measurement" OR "measuring performance" OR "performance measure" OR "performance measures"[1].

The search resulted in the identification of 2,310 citations as of November 13th, 2009. Citations were imported into an Excel Spreadsheet, EndNote Bibliographic Management Software, and Sitkis Access database.

The effort generated 48,798 cited references for the 2,310 citations. This large data clearly required to be adequately screened and cleaned. Data was exported into Excel Sheets and two key tables were created for the cited articles and the citing articles. This study acknowledges concerns that surround citation analysis as identified in the literature (Pilkington and Liston-Heyes, 1999). In particular, many duplicates were found within the cited reference data due to variations in author

---

[1] The finalized search terms (July 2009) were:
    TS=(("performance measurement" OR "measuring performance" OR "performance measure" OR "performance measures"))
    Refined by: Subject Areas=(OPERATIONS RESEARCH & MANAGEMENT SCIENCE OR MANAGEMENT OR BUSINESS)
     [excluding] Document Type=(EDITORIAL MATERIAL OR NOTE OR BOOK REVIEW OR CORRECTION OR REPRINT OR LETTER)
    NOT ts="mutual fund" NOT ts="mutual funds" NOT ts="capital gains" NOT ts="bond fund" NOT ts="options backdating" NOT ts="stock option" NOT ts="stock options" NOT ts="active trading"
    NOT Conference Proceedings

names and to missing or inaccurate information in the cited references themselves. Indeed, problems with data quality within cited references in ISI Web of Science are well documented in the literature (Clark, 2008, pp. 9, 19, 21; Gibbs and Sargeant, 2009, pp. 7, 8). However, the careful review of the data as well as the large size of the sample size can be considered as measures and factors that ensured the validity of this citation analysis.

Data produced by the citation analysis was analyzed by the researcher using Access. Three key queries were run to identify the (1) top most cited works over the examined period, (2) the number of publication per year, and (3) the top five most cited first authors every year. Next, the researcher provided a colleague who has strong Access skills with the citation analysis dataset and asked her to independently run the same queries. The results of the two separate queries were similar but both Access analyses underscored a problem with duplicate entries. The query about most cited works pulled up two Harvard Business Review works for Kaplan for the same year, one of which with no detailed references.

The researcher then asked another librarian colleague with experience in citation analysis to cross check the results of the top cited works query with the datasets from the citation analysis. Both the researcher and the librarian determined with confidence that the two Kaplan's entries were a duplicate and one of the entries was consequently removed from the final results table. As a precautionary measure, the researcher asked the librarian to examine the results of the other two queries to ensure there were no duplicates in that data. Upon verifying that no duplicate entries existed, the results of the Access analyses were put into final tables that are discussed in the following section.

1.2.2   The state of performance measurement research field: key findings

Performance measurement is regarded as an emerging academic sub-discipline that falls within the discipline of management (Thorpe and Beasley, 2004). However,

academic research in the field of performance measurement is believed to originate from different disciplines and functions, including operations management, economics, accounting, marketing, human resources, psychology, and sociology (Marr and Schiuma, 2002).

Citation analysis was used in this research to explore the growth of the field of performance measurement among researchers and add to our knowledge about the main contributors to the field over the past twenty nine years. It was also used to identify the most prominent papers or works and look into the extent of their impact on the field. An examination of the frequency of citations for individual pieces of work can help identify the works that have been most influential in the field. This examination can also improve our understanding of the theoretical and conceptual ideas that constitute the major foundations of the performance measurement field. Table 1.1 displays the top ten most cited works in the field of performance measurement between 1980 and November 2009.

The first observation was the heightened attention in the field to a rather limited number of works and a noteworthy variation between the citation numbers for each of the top ten most cited works. For example, the fist two most cited work of Charnes et al and Kaplan & Norton were cited 98 and 93 times respectively. After that, the number of citations drops significantly until it reaches 44 times for the tenth most cited work. In fact, each of the last four most cited works in the field was cited 50% less than the 1978 work of Charnes et al. Similarly, each of the last three most cited works in the field was cited 50% less than the 1992 work of Kaplan and Norton. This is a key indicator of the relative weight and importance of the top two works listed in table 1.1 compared with the rest of the list.

Moreover, the top five most cited works, which included two different entries for Kaplan and Norton, were cited 364 times collectively compared with 234 times for the collective five works that follow. This means that the top five works of Charnes, Kaplan, Neuts, and Dixon were cited 55.6% (130 citations) more than the five works that follow as listed in table 1.1.

Perhaps the most prominent feature about the data displayed in table 1.1 was the dominance of the work of Robert Kaplan, as a first author of a most cited work, in the field. The two 1992 and 1996 works of Robert Kaplan were cited 145 times in total. This number of citation heaps to 189 out of a total of 598 citations for the top ten works, if the 1987 work of Johnson and Kaplan was added. This means that works that belonged to Kaplan, as a first and second author, were cited more than 31% out of the total number of citations for the top ten most cited works in the field.

Another noteworthy observation was related to the relatively old era of most of the top cited works. In a citation analysis that included publications until November 2009, it was interesting to find out that the most cited work in the field was 31 years old. The trend continued to persist throughout the list with the second most cited work being 17 years old and the third most cited work being 28 years old at the time of conducting the citation analysis. This observation should be approached with care as citation analysis tends to naturally favor older publications. However, the relatively old age of the top works in the field raised legitimate concerns about lack of continuous supply of important contributions to the field. It also triggered the question of whether the field has reached its limit as far as producing new knowledge and worthy contributions.

The journals from which citations appeared were also spread and included the European Journal of Operational Research, the Harvard Business Review, Management Science, Journal of Management, and the International Journal of Operations and Production Management.

Work	Reference	Author(s)	Year	Citations
'Measuring the Efficiency of Decision Making Units'	European Journal of Operational Research	Charnes, A., W. W. Cooper, and E. Rhodes	1978	98
'The Balanced Scorecard - Measures That Drive Performance'	Harvard Business Review, Jan-Feb	Kaplan, R. S. and D. P. Norton	1992	93

Matrix Geometric Solutions in Stochastic Models : An Algorithmic Approach	Baltimore, MD: John Hopkins University Press	Neuts, Marcel F.	1981	62
The New Performance Challenge: Measuring Operations for World-Class Competition	Irwin Professional Publication	Dixon, J.R., A.J. Nanni, and T.E. Vollmann	1990	59
The Balanced Scorecard: Translating Strategy into Action	Boston, Mass.: Harvard Business School Press	Kaplan R.S. and D. P. Norton	1996	52
Stochastic Models of Manufacturing Systems	Englewood Cliffs, N.J.: Prentice Hall	Buzacott, John A. and J. George Shanthikumar	1993	51
'Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis'	Management Science	Banker, R. D., A. Charnes, and W. W. Cooper	1984	48
'Firm Resources and Sustained Competitive Advantage'	Journal of Management	Barney J.	1991	46
'Designing, Implementing and Updating Performance Measurement Systems'	International Journal of Operations & Production Management	Bourne, M., J. Mills, M. Wilcox, A. Neely, and K. Platts	2000	45
Relevance Lost: The Rise and Fall of Management Accounting	Boston, MA: Harvard Business School Press	Johnson H.T. and R.S. Kaplan	1987	44

Table 1.1: Ten Most Frequently Cited Performance Measurement Works

An examination of the number of publications per year aimed at identifying the growth or decline trends in the field. The results of the examination showed that publications in the field of performance measurement have been steadily increasing since 1990 despite some dips at certain years. The number of publications as of November 2009 has reached 186 compared with 104 publications in 1999. This result can probably be accepted as an indicator of a healthy and growing field, at least in terms of its sheer numbers of publications.
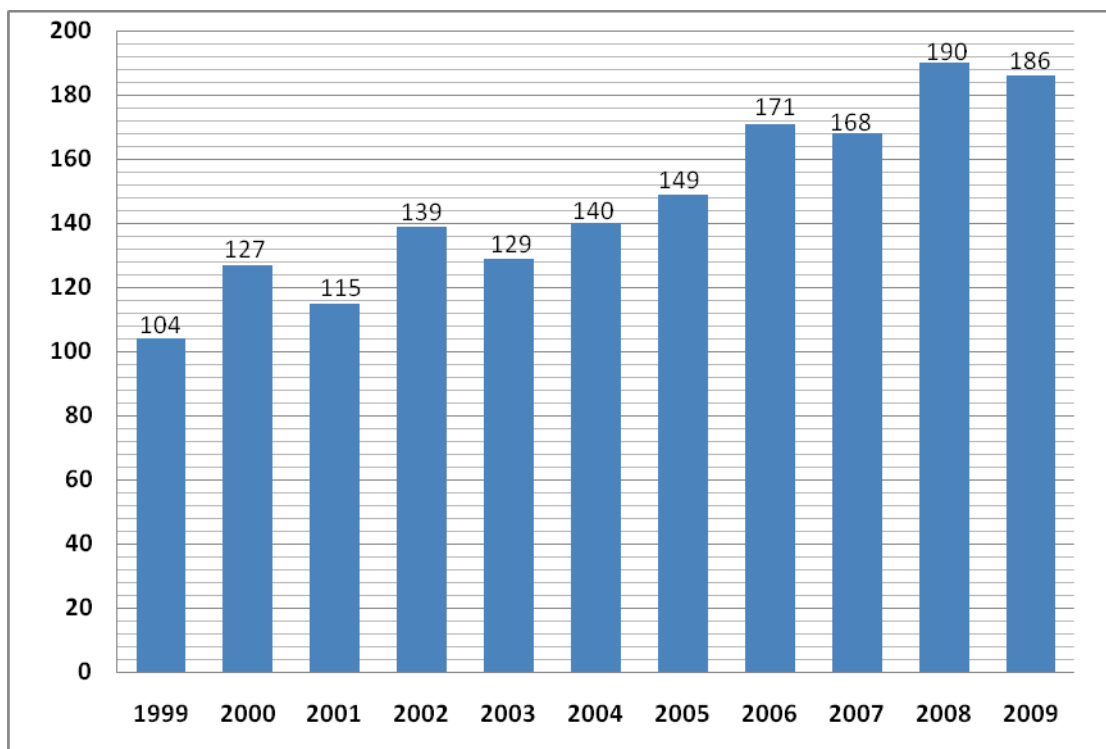
Figure 1.1: Performance Measurement Publications per Year (1995-2009)

An annual count of citation frequencies (Table 1.2) revealed a remarkable observation about a monopoly of authors in the field. Robert Kaplan has been the most cited first author every single year for the last ten years. Moreover, the annual counts showed that for most years there was a significant gap in the number of citations between Kaplan and the second most cited author for that year. In more recent years, (2001, 2006, 2007, and 2009), Kaplan has been cited twice or as much

as the second most cited author. The monopoly seems to be taking root at an increasing rate.

In four out of the last five years (2005, 2007, 2008, and 2009), Andy Neely has been the second most cited author falling immediately after Robert Kaplan. Abraham Charnes came next after Andy Neely as a second most cited author.

The data also revealed another author who was third most cited author in more than one year as of late. This was Umit Butitci who was the third most cited author in 2005 and again in 2009.

Year	1st Most Cited	2nd Most Cited	3rd Most Cited
2009	Kaplan R.S. (91)	Neely A. (40)	Bititci U.S. (22)
2008	Kaplan R.S. (62)	Neely A. (36)	Banker R.D. (22) Saaty T.L. (22)
2007	Kaplan R.S. (80)	Neely A. (32)	Porter M.E. (19)
2006	Kaplan R.S. (74)	Artalejo J.R. (26)	Neely A. (24)
2005	Kaplan R.S. (82)	Neely A. (53)	Bititci U.S. (28)
2004	Kaplan R.S. (41)	Whitt W. (24)	Neuts M.F. (16) Charnes A. (16) Ehrenberg A.S.C. (16)
2003	Kaplan R.S. (28)	Kanji G.K. (22)	Mentzer J.T. (18)
2002	Kaplan R.S. (43)	Neely A. (24)	Cooper R.G. (18)
2001	Kaplan R.S. (37)	Hayes R.H. (18)	Neely A. (17) Lee C.Y. (17)
2000	Kaplan R.S. (33)	Charnes A. (21)	Neely A. (16)
1999	Kaplan R.S. (30)	Charnes A. (21)	Whitt W. (15)

Table 1.2: Citation Frequencies Annual Count Over the Past Ten Years

Moreover, the results of the annual count of citation frequencies were interesting from a performance measurement 'discipline' point of view. Robert Kaplan has an

accounting disciplinary background. Andy Neely and Umit Butitci have an operations research disciplinary background; Abraham Charnes has a mathematics and operations research background.

While most of the key contributors to the field have operational research disciplinary background, the most prominent contributor, Robert Kaplan, has an accounting disciplinary background. If the field of operational research had indeed made major contributions to the debate about performance measurement at both strategic and operational levels (Wisniewski and Shafti, 2003), then why has performance measurement research been so strongly dominated by a researcher from the field of accounting. Perhaps the financial accounting 'roots' of performance measurement could have explained dominance by researchers from the accounting field in the early years of performance measurement. The discussion in chapter 2 highlights how the field started with a strictly financial focus. However, performance measurement has supposedly matured and developed into a more encompassing field as the discussion in the next chapter indicates.

The question then becomes why has this 'alleged' evolution not been reflected by strong contributions from other fields. This is further complicated by concerns about a widespread of the Balanced Scorecard, Kaplan's work, which is not matched by research into the implications of adopting this framework. Researchers continue to warn that, despite its widespread use, there is still a strong need to investigate the Balance Scorecard as a strategy for performance management and measurement, especially in developing countries (Gomes and Liddle, 2009).

1.3   Structure of the Research Study

This chapter offered a general introduction to the topic of organizational performance measurement and highlighted its key features and characteristics. The impetus for conducting the study was presented along with the general reasons for selecting to focus on the topic of measurement data utilization. The chapter also

14

presented the results of an extensive citation analysis that examined publications in the field from the 1980 till the end of 2009. The results confirmed assertions in the literature about an increasingly growing field. Results also highlighted a strong dominance of one approach to performance measurement as underlying theory for the field and revealed a monopoly of one author in the field of performance measurement.

Chapter two tracks the historical threads of performance measurement in the public sector around the world. The reasons behind the shift in the focus of performance measurement are presented to offer a better understanding of the background and context within which a new performance measurement paradigm was born. Several definitions of performance measurement are also presented along with the alleged roles and functions that these systems can play in a public sector setting. Three performance measurement frameworks are selected as examples and discussed with some level of detail to provide a better sense of what a real life performance measurement system entails. A summary of the key challenges that typically accompany the adoption and implementation of performance measurement systems are presented including: problems with designs, quantity of performance measures, readiness of organizations, and misuse of measurement data. The chapter ends with a clear identification of the gap in the literature that this research study will focus on.

Chapter three starts with stating the research questions that guide the research effort to follow. The traditional and emerging trends in management research are discussed before presenting four main paradigms to research methodology and highlighting the main characteristics of each. The discussion about the four paradigms (realism, constructivism, positivism and post positivism, and critical theory) helps set the selected method for this research in perspective to the different available research paradigms. The 'case study' method is presented and discussed in details as the selected strategy for this research. The quality tests of validity and reliability are discussed as they pertain to case study research. The aptness of using personal in-depth interviews and semi-structured questions in this study is briefly

presented prior to a full discussion in the chapter that follows about data collection methods.

Chapter four includes a detailed account of all the secondary and primary data collection processes. A description of the case study organization and the education reform provides the necessary information about the background and context of the case study. The tenets behind establishing the performance measurement system at the case study, and the components of this system, are also described in full details. An intensive effort to review over 1000 media records is described and the results of the review inform this study by shedding light on the reactions of the society to the reform in general and the new performance measurement in particular. For the primary data collection effort, the careful process of identifying key stakeholders is presented along with details about conducting the personal interviews and the focus group. The process that was followed during the data analysis stage is also described in the chapter including: a full account of the different steps of data display, reduction, and conclusion drawing; the coding sheets and the ways these were used and filled; and the themes that emerged during the data analysis.

Chapter five presents five key findings from this research study. The findings are drawn from a combination of tested priori codes as well as emergent themes that surfaced during the data analysis stage. Not all of this research' findings are consistent with what the literature asserts as factors that influence the utilization of performance measurement data. Moreover, a number of novel attributes and factors are found to influence the utilization of measurement data in this case study.

Chapter six starts with providing clear answers to all the research questions listed in chapter three. Fifteen factors that are found to influence the utilization of measurement data are presented as part of the discussion about key conclusions drawn from this study. This is followed by suggesting five major recommendations to promote performance measurement systems and improve the utilization of their outcome data. Next, the chapter lists six significant contributions that this research made to theory and practice. A critical assessment of the quality and limitations of

this study aims at demonstrating the validity of this research and the reliability of its process. The chapter concludes with suggesting several areas for future research.

CHAPTER 2

LITERATURE REVIEW

2.1   Introduction

Performance measurement is a topic that has caught the attention of a variety of audience for decades now (Holzer and Kloby, 2005). Indeed claims for the importance of performance measurement systems have been made in the literature for decades; books and journal articles on performance measurement and management abound (Kaplan, 1984; Faucett and Kleiner, 1994; Neely *et al*., 1995; Bourne *et al*., 2000; Bourne, 2002; Gianakis, 2002, Meyer, 2002; Courty and Marschke, 2003; Johnsen, 2005; Greiling, 2005). Many public administrators and academics clearly believe that performance measurement systems can potentially fulfill the promise of professional public-sector management (Gianakis, 2002). Around the world reports abound about elected and politically appointed officials who consider performance measurement an objective tool capable of both improving decision making and fostering fiscal prudence (Holzer and Kloby, 2005).

Furthermore, performance measurement systems constitute a central element of what has come to be known as the New Public Management (Halachmi and Bouckaert, 1996); a doctrine that originated in New Zealand, the United Kingdom, and Australia and was later on transferred to the United States (Moynihan, 2006). The ideas of this doctrine, a set of arguments about how governments should be run, were well received by an international audience eager to reform its public sector (Osborne and Gaefler, 1992). Public sector performance measurement systems developed into a phenomenon that is indeed international in scope (Holzer and Kloby, 2005).

Some arguments for the significance of performance measurement systems go as far as posing the important functions that performance measures play in political competitions as analogous to prices in market competition: "*PIs may effectively function as carriers of information, functioning as 'prices' in political markets, in much the same way as prices do in input and product markets. This can explain why different stakeholders consistently screen performance reports for information that they can use strategically in politics. Selected PIs may thus effectively create 'creative destruction' of the present or future political or managerial status quo.*" (Johnsen, 2005: 14-15). This is a serious argument that, at least on the face of it, entails strong support to performance measurement systems. However, the point about the power that Performance Indicators (PIs) are likely to hold warrants further discussion and will be addressed later on in the Chapter under the challenges associated with measurement systems section.

The critical thrust of the result-oriented performance measurement promise remains in its aptitude to monitor policy goals and hold public managers accountable for policy outcomes (Johnsen, 2005; Gianakis, 2002). Consequently, a shift from the traditional way of ensuring accountability through budgetary control allows for a new paradigm where public managers can more effectively utilize their professional expertise and good judgment to realize the desired outcomes. Performance measurement systems can hence be viewed as a fundamental tool to resolve the typical predicament of effectiveness versus control that faces reform in public sector management (Moynihan, 2006): "*the simultaneous desire for professional and efficient administration and democratically accountable government* "(Gianakis, 2002).

While the present status of the performance measurement field in the public sector can be epitomized by "*lively debates that take place where 'true believers', 'pragmatic skeptics' and 'active doubters' are engaged simultaneously*" (Norman, 2002), a review of performance measurement definitions and roles is warranted to shed light on the scope of the topic before attempting to understand its 'pros' and 'cons'.

2.2   Historical Background in the Public Sector

A good starting point to better understand performance measurement systems can be achieved when the historical threads that trace the evolution of these systems are captured. A historical review of performance measurement systems indicates that performance measurement has a long tradition within public policy and management. The first era of performance measurement in the public sector, or some elements of it, can be traced as far back as the beginning of the 20th century through the efforts of the New York Bureau of Municipal Research starting around 1912 (for the history of performance measurement systems see: Bouckaert, 1992; Gianakis, 2002; Williams, 2003; Halachmi, 2005; Holzer and Kloby, 2005). Bureau leaders redefined the budget from a tool for accounting to an instrument capable of showing the performance of government, fostering transparency, and holding elected officials accountable for their tasks (Williams, 2003).

Invigorating citizen trust, confidence in leaders, and active participation was thought to be accomplished through educating the public as to the budget process and providing material via accounting publicity (Holzer and Kloby, 2005). By 1930, performance measurement, and in particular performance budgeting, had progressed into an advanced management tool and served as a useful political device for observing government at work. The early practices of line item control budgeting were succeeded later by a decision focused budget.

Years later, in 1950, the first performance budget was created based on the recommendation of the Hoover Commission Report. Performance measures slowly emerged as a tool for efficiency, enabling mayors, city managers, governors, or expert administrators to get results with limited resources. This happened because at the time performance measures focused on efficiently using financial inputs to yield maximum results. The period of 1950s and 1960s reflects the rise of performance budgeting; "management by remote control", - i.e. managing by the financials figures, became a popular management tool (Johnson, 1992). This period also

reflects the rise of post-war generation of statistical social scientists in the United States and the expansion of program evaluation; largely the same thing as performance measurement but practiced by social scientists with a different set of skills than public administrators (Williams, 2004).

During the 1970s a new doctrine was being introduced and established in the United Kingdom, New Zealand, and Australia that comprised of new ideas about how governments should be run. The doctrine, which was later on dubbed as New Public Management reforms (NPM), revived interest in performance measurement systems as reflected in the body of research. NPM was transferred to the United States towards the end of the 1970s and the beginning of the 1980s (Halachmi and Bouckaert, 1996). Regarded by some authors as a global movement that reflects liberation management and market-driven management, NPM is epitomized by having performance measurement as a central element (Gianakis, 2002). Liberation management is concerned with relieving public managers from unnecessary rules and regulations and shifting their control and focus from input factors to outcome measures. Market driven management requires performance measures allowing organizations to compete for budgetary resources (Greiling, 2005). The 1970s and 1980s also witnessed the popularity of a number of systems that basically followed the general guidelines of NPM. Zero-based budgeting and management-by-objectives are examples of such systems where program outcome measures and performance targets were used respectively (Greiling, 2005). The early 1980s also witnessed the introduction of the Financial Management Initiative in the UK which embodied performance management as a central piece of the initiative (Propper and Wilson, 2003).

Concepts of public performance measurement were once again boosted as they received a further push to the center of attention through the work of Osborne and Gaebler (1992), the work of the associated National Performance Review (1993), and the introduction of the Government Performance and Results Act (GPRA) in 1993. The report submitted to the Senate Committee suggested a declining public confidence in the institutional effectiveness of American government. Public

opinion polls revealed that this perception was particularly true for institutions at the federal level of government. Increased public desire for a wide range of government services was complemented by increased public disregard for government and opposition to paying higher taxes. The committee partially attributed this inconsistency to the public's dissatisfaction with the level and quality of government services in exchange for their tax dollars. Plans to cut red tape, hold citizens first, empower employees to get results, and produce better government for less were introduced by the National Performance Review (Gore, 1993). Moreover, the introduction of the GPRA provided for the establishment, testing, and evaluation of strategic planning and performance measurement in the American federal government. According to this legislation, federal institutions are obliged to establish strategic plans with unambiguous objectives. Progress on these objectives is monitored and evaluated using performance measures. Some of the foremost goals of the GPRA are to improve public confidence, federal management, program effectiveness and public accountability by promoting a new focus on results, service quality, and citizen satisfaction (Gore, 1993).

While US efforts were designed to foster improved fiduciary decision-making via objective information, member states throughout the OECD were also producing a new model of public governance incorporating a more modest role of the state as direct public service provider and a strong role of performance measurement (Sanderson, 2001). Several countries have paved the way in this development. Most notably, the UK has introduced a series of performance measurement initiatives to improve the quality of the public sector. Under the Conservative governments of the 1980s and 1990s performance monitoring was developed by the administration to ensure that public services were providing value for money. In fact it was not until the introduction of organizational and managerial reforms introduced by these conservative governments that public sector performance measurement became firmly established (Boland and Fowler, 2000). In particular, both the education and health sectors witnessed widespread adoption of performance measurement systems. The education sector adopted performance measurement systems through establishing the Office for Standards in Education (OFSTED) in 1992, and

introducing Performance Indicators (PIs) to collect and publicize summary information including truancy rates and GCSE pass rates. The health sector adopted the Performance Assessment Framework (PAF) to measure its organizational performance (Chang and Northcott, 2002). Following the election of the labor administration in 1997, there has been a notably large increase in the use of performance measures in the UK (Propper and Wilson, 2003). Over the last decade, the National Audit Office (NAO) has transformed value for money traditionally focusing on financial audit into new approaches to performance audit that typically involved complex delivery systems of public services, many stakeholders and long time horizons (Ling, 2002). It must be noted here that ideas and prescriptions of NPM are reported to have been applied in the UK more consistently than in many other countries over a quarter of a century (Cutler and Waine, 2005). This might explain why a large portion of performance measurement literature comes from the UK since this field constitutes a central aspect of NPM reforms.

Elsewhere in Europe, the Cabinet's white paper 'Van Beleidsbegroting tot Beleidsverantwoording' (VBTB; "From policy budgeting to policy accountability") was issued in 1999 containing the outline for budgeting and accountability reform in the Netherlands. It spearheaded the ambition to transform existing resource-oriented budgets into policy-oriented budgets with a more explicit connection between resources, performance and objectives. Aiming at an improvement of the information value and accessibility of national government budgeting, since 2002, all budget documentation and annual reports must be developed in the new 'VBTB' style (Ministerie van Financiën 2002). Other countries, front running in performance measurement, include Australia, New Zealand, and the Nordic European countries.

2.3   The Shift to a Public Sector Unique Paradigm

The above discussion about the history of performance measurement systems in the public sector serves to illustrate that initially, and similar to its constituents in the

private sector, the most commonly used performance measurement systems focused on efficiency; the extent to which the public organization's resources have been used in the shrewdest manner. Efficiency in this context is defined as the relationship between outputs and inputs. Similarly, performance measurement in the private sector has traditionally been of a financial nature (Neely, 1999; Halachmi, 2005; Holzer and Kloby, 2005). Accounting frameworks have existed since the Middle Ages in the private sector, and by the start of the twentieth century, with the increasing separation of company management from ownership, the need for owners to monitor managers resulted in the use of such measures as return on investment to judge managers' performance (Johnson, 1983). These systems continued to develop and advance in the private sectors across the years where interest in performance measurement systems started to move away from being efficiency-centered and became mainly linked to the effectiveness of strategic management systems as well as narrowing the gap between ambitious strategies and operational plans (Greiling, 2005).

While the private sector paved the way for using performance measurement systems in arenas outside pure efficiency, measuring organizations' performance in the public sector seemed to have moved at a rather slower phase, followed quite a different path, and only recently has it gained momentum and secured the necessary legislative support in some countries as demonstrated in the discussion above. The different path of progression that performance measurement systems followed in the public sector was necessitated by the special characteristics of this sector. The very notion of measuring the performance of organizations in the public sector was often deemed contentious and considered to be difficult because of the very different nature of these organizations. Frequently quoted reasons by skeptics who believe performance measurement systems are problematic, if not unsuitable, to the public sector include low profit maximizing interest, little potential for income generation, and, essentially, no bottom line to ultimately measure performance against (Boland and Fowler, 2000).

Moreover while public organizations theoretically only have one shareholder, the State, they do have to account to a range of stakeholders including users of their services, payers for their services, and professional organizations. They can also be held accountable to politicians and legislative bodies that in essence represent millions of citizens. The notion of having to serve several masters, (Dixit, 2002), means that a public organization often has multiple principals to satisfy and, as a result, multiple tasks to undertake. The feature of multiple principals and multiple tasks means that the goals of a public organization might sometimes be vague or even in conflict (Propper and Wilson, 2003). This adds to the complications that are likely to surface when using performance measures to evaluate the performance of public organizations. In short, the multiple, and at times blurred, goals of the public sector make it difficult to measure the performance of these conflicting and unclear goals.

Nonetheless, and similar to their private sector counterparts, public organizations have grown increasingly interested in managing and measuring performance. One of the reasons behind this interest is the continuing pressure for accountability and value for money. As public organizations in most countries are spending tax money citizens have the right to be assured that it is spent wisely. Another reason is the continuing pressure on government budgets which forces public organizations to monitor and evaluate how to be more efficient.

Additionally, contemporary public management represents an entrepreneurial, market-oriented way of viewing the government, in which management increasingly relies on rational economic methods such as accounting and performance models (Harris, 1998). Finally, increasing external pressures on public services from international agreements and trans-national bodies, particularly in Europe, have increased the demand for and requirements of performance audits (Ling, 2002). Many authors have observed these changes in the public sector in an international context (Pollitt, 2000; OECD, 2000; Ling, 2002; Mills *et al.*, 2002; Walsh, 1995).

The discussion above clearly indicates that performance measurement systems in the public sector had to adapt to, and embrace, a far more complex set of unique factors. Issues related to improving external accountability and increasing the efficiency of internal processes constituted prominent priorities since the early days of introducing performance measurement systems to the public sector (Greiling, 2005).

Moreover, performance measurement systems were increasingly expected to play an even larger role in modernizing public management especially in the aftermath of the NPM doctrine. Advocates of NPM ascribe a high priority to measuring outputs and outcomes and believe that management activities and new policies must be based on this type of information (Van Thiel and Leeuw, 2002). In essence, the resulting measures are now expected to act as the "new" monitoring mechanism that enable politicians to measure and evaluate the performance of public organizations, monitor policy goals, and achieve accountability. Performance measures became an essential part of the new public sector management equation.

Against this backdrop of changes in the nature of the public sector and the evolving expectations of new measurement systems, many governments adopted policies and measures to institutionalize, systemize, and regulate the availability and use of information produced by theses systems. Access to measurement information can arguably be considered critical to facilitate the 'new' monitoring role of the measures and address principles of transparency and accountability. This perhaps explains why most views of NPM highlight its critical relationship with concepts of reporting; some even use the two terms of performance measurement and performance reporting interchangeably.

Several forms of external requirements, such as mandated performance reporting, were increasingly introduced by several governments around the world rallying to adopt new ways to manage their public organizations. The previously referenced GPRA is one example of an external requirement. However, the formulation of policy requirements has not been proven yet to be the solution to measurement data availability and criticism abounds regarding the value and effect of mandated

external reporting (Holzer and Gabrielian, 1998). Some findings even indicate that internal reporting requirements, mandated by a public organization's management, are more efficient than externally mandated policies (de Lancer Julnes and Holzer, 2001).

Indeed, the inadequacy of traditional, financially focused, performance measurement systems to serve all these new roles and purposes and meet the requirements of a modern public sector soon became apparent. In particular, the 1980s and the 1990s witnessed accelerated criticism of traditional measures during the 1980s and early 1990s (for example see Johnson and Kaplan, 1987; Hiromoto, 1988; Miller and Vollmann, 1985). Ironically, many of the most vocal critics of traditional performance measurement systems have come from the academic accounting community (Neely, 1999). A number of reasons were cited in the argument against solely relying on financial information in decision making. Challenges related to financial measures can be summarized as (1) providing little guidance on future performance, (2) encouraging short termism, (3) internally rather than externally focused with little regard for competitors or consumers, (4) lack strategic focus, (5) often inhibit innovation, and (6) provide little indication of how performance is achieved or how it can be improved (Kennerley and Neely, 2002).

The realization that accounting systems alone were not sufficient anymore placed the need for more "balanced" performance measurement systems on the management agenda and sparked intensive research efforts in the academic community. A more balanced approach was seen as one that is more inclusive of intangible assets, stakeholders and the environment an organization operates in. Arguments supporting the need for these more balanced approaches to measuring accountability and performance in the public sector gradually gained ground over the past few decades. By the late eighties and early nineties, the literature witnessed a thriving dialogue on the subject. In 1992, the balanced scorecard (BSC) approach was the first attempt to address the shortcomings of traditional, single-focused, performance measurement systems (Kaplan and Norton, 1992). The balanced scorecard, and other comprehensive approaches, is discussed in the next section. The discussion

also highlights the role that the ideas and principles of Total Quality Management played into the development of theses more comprehensive approaches.

In summary, the evolution of performance measurement systems in the public sector has witnessed a gradual acceptance of and shift towards more comprehensive approaches. This change was further reinforced in certain countries by legislative requirements that call for a more balanced approach in measurement and encourages the inclusion of non-financial performance measures (for example: the GPRA in the USA, linking performance targets to resources allocated by the Treasury to government departments in the UK). As a result of these changes, a myriad of performance indicators surfaced to measure quality of provided service, satisfaction of clients, efficiency and effectiveness of internal business processes as well as capacity and satisfaction of internal human resources. The end basket of selected performance measures serves as a control system to ensure that the total performance of an organization is effective, on track, and yielding the aspired results.

2.4   Definitions and Roles of Performance Measurement Systems in the Public Sector

An endeavor to identify and land on a clear, unified and agreed upon definition of performance measurement systems in the literature reveals an almost mercurial nature of a field that is further riddled with questions related to roles, scopes, and frameworks of operation. It is possible that this is related to the multi-faceted nature of organizational performance as a concept; one that inherently involves many disciplines such as operations management, accounting, human resources, business strategy, and organization behavior (Neely, 1999). Box 2.1 provides few examples of some definitions found in the literature illustrating the varying orientations that underpin the field and depicting the wide range of differences in its scope and focus:

A performance measurement system is…

… A process of defining a mission and desired outcomes, setting performance standards, linking budget to performance, reporting results, and holding public officials accountable for those results. (The National Academy of Public Administration, 1998)

… The ongoing monitoring and reporting of program accomplishments, particularly progress towards pre-established goals. (US General Accounting Office, 1998)

… Managing and reporting based on what programs are achieving for citizens and at what cost. This implies agreeing on expected outcomes, measuring progress toward them and using that information to improve performance and report results. (The Office of the Auditor General of Canada, 1998)

… Using performance information effectively and performance measurement is obtaining the requisite information. (KPMG, cited in English, 1998)

Box 2.1: Definitions of Performance Measurement Systems

(Source: Davies, 1999)

Moreover, the literature utilizes a number of terms that are sometimes used interchangeably to describe different components of a performance measurement system. The proliferation of definitions in the literature results in a degree of perplexity regarding, in particular, the differences between performance measurement, performance measures and performance measurement systems. In order to distinguish between the three concepts, performance measurement can be defined as the processes of quantifying the efficiency and effectiveness of an action (Neely *et al*., 1995). In other words, performance measurement can be described as the "the development of indicators and collection of data to describe, report on and analyze performance" (Marshall *et al*., 1999). On the other hand, a performance measure, also known as performance indicator, can be defined as a metric that is

used to quantify the efficiency or effectiveness of a particular action (Neely *et al*., 1995) as well as measure the quality of that action. A performance measure can take the form of a number, percentage, ratio or scale. A performance measurement system is the set of metrics that are used to quantify both the efficiency and effectiveness of an action (Neely *et al*., 1995). In other words, it can be regarded as the overall umbrella under which performance measures are placed and performance measuring takes place. This means that a performance measurement system can be regarded as the summation of a select number of performance measures that are deemed significant and indicative of the efficiency, effectiveness and quality of major actions in an organization. Performance measures are components of a larger performance measurement system.

The above discussion about definitions introduce performance measurement systems as management tools that attempt to measure the efficiency, effectiveness, and quality of actions taken by an organization. They are tools that are developed to improve a variety of management functions such as resource allocation, business process formulation, improving accountability, inter –administrative comparison and bench marking, performance monitoring, and internal diagnosis and decision making processes. The uniqueness of performance measurement systems rests in their potential capacity to provide an organization with a continuous flow of data that can cover most, if not all, aspects of its operation. The collected data can potentially provide "objective" feedback that allows managers to evaluate the impact of their actions, examine performance and monitor progress. An organization needs to go through a process of defining a number of key tasks and activities; in turn this allows for performance measures to be set and reported against in order to monitor progress (Harborne, 1999).

It also follows that, as management tools, performance measurement systems have the potential to play a critical role in guiding the planning function at an organization. The benefits of using real data provide a strong argument in favor of using outcomes of performance measurement systems to make more grounded judgment about past performance and, as a result, provide more objective guidance

to the planning process. This means that performance measures has the potential to take much of the guess work out of the planning process, since a measured record of past performance is likely to constitute the best indicator of future performance (Harbor, 1997). In cases when planning calls for a given result or process to be kept within a certain ranges or levels of performance, performance measures can be used to articulate these ranges or levels as well as to monitor adherence to the plan in place. This has important implications on the nature of the relationship between performance measurement systems and planning as a management activity if indeed outcomes of the planning process do in turn affect the design and focus of a performance measurement system.

Since each of the two concepts, performance measurement systems and planning, seem to have a great effect on the other it may be supposed that measuring the performance of an organization is likely to affect its strategic planning and strategic decision making. The provided feedback from performance measurement systems can reshape or change strategy. At the same time, and for selected performance measures to serve their purposes, performance measurement systems need to be aligned with the set strategy of an organization and closely related to the goals and objectives that cascaded from this strategy. This link is essential if an organization is serious about its proclaimed strategy and is keen on adopting an integrated approach that binds long term plans with day-to-day operations.

Hence one of the key purposes of performance measures is to align the entire organization behind the goal of turning the strategic plan into effective action (Aguilar, 2003). Performance measurement systems can then be viewed as a key part of strategic control and have the capacity to exercise control through: helping managers to identify good performance; setting targets; and demonstrating success or failure (Neely *et al*., 1994). Consequently it can be stated that performance measurement systems can help manage the organization through setting the strategic direction and goals, monitoring and controlling the implementation of strategies, linking strategy to operation, and providing feedback that can reshape or change

strategy. In all of these roles, performance measurement systems support better decision-making and enhance organizational learning and growth.

Other roles that performance measurement systems play in the public sector include modernizing public budgeting systems and introducing obligatory or voluntary reporting (Gianakis, 2002; Greiling, 2005). Shifts from input based to output/outcome based public budgeting are increasingly considered a key element of public sector reforms. In that respect, performance measures offer the means to provide organizations with output/outcome indicators that can be included in the budget and planning process (Greiling, 2005).

On the other hand, reporting (internal and external), can be considered another important element of public sector reform initiatives. Providing information about how public organizations are fulfilling their mandates and the quality of their services is essential to achieving greater transparency and accountability. Improving public accountability is related to fulfilling the old desire for professional and efficient administration as well as democratic and accountable government (Gianakis, 2002). In that regards, performance measurement systems can play an important role in providing organizations with the right mix of indicators, and as a result performance measurement information, that can be used for reporting purposes.

2.5   Performance Measurement Frameworks

Performance measurement systems in the public sector had to adapt to a number of changing conditions and requirements over the years. New Public Management calls for comprehensive and innovative approaches that transcend the strictly financial focus of traditional models and address the different roles and expectations of the public sector. A number of performance measurement frameworks were developed over the past few decades to meet the new requirements and cater for an increasingly complex set of factors in the modern public sector. For purposes of providing a

better idea about these systems and illustrating some of their features and components, three frameworks are selected and elaborated upon. It is important to note that the selection is based on the literature review as well as the researcher's working experience. At least two of the selected frameworks are noted by the researcher to be frequently used by consultants and development agencies such as the United Nations and the World Bank. These are the Balance Scorecard and the Logical Model. As discussed in the first chapter, the Balance Scorecard is also heavily referenced in the literature and considered to be one of the most popular performance measurement frameworks. The literature also makes reference to the Performance Prism as a performance measurement framework that is specifically designed for the public sector. The researcher thought it would be particularly relevant for this study to select and illustrate a framework that is oriented towards the goals and purposes of public organizations.

The following section discusses these three frameworks: the balanced scorecard, the performance prism, and the logic model.

2.5.1   Balanced Score card (BSC)

The Balanced Score card (BSC) represents one of the first attempts to address some of the challenges associated with financially focused performance measurement systems. It was introduced in 1992 by Kaplan and Norton as a comprehensive framework claiming to address the shortages of single-focused approaches. In tracing the historical origins of the BSC framework it becomes clear that concepts of Total Quality Management have strongly influenced the development of this framework. The original ideas of Edward Deming, emphasizing that quality improvement is achieved through the statistical control of processes and the reduction of variability of these processes (Munro-Faure and Munro-Faune, 1992), were further expanded by Dr. Joseph Juran who broadened quality from its original statistical origin to total quality management and control (Ishikawa, 1985). These ideas were later on developed by Art Schneiderrman who developed what is known as the first Scorecard with goals for a series of quality measures that correspond to

what he considered to be critical success factors for the company he worked for at that time (Anthony and Govindarajan, 1997). Schneiderman was later on part of the Nolan-Kaplan group on performance measurement that led to the development of the Kaplan-Norton BSC (Schneiderman, 2001).

The BSC framework identifies and integrates four different ways of looking at performance (financial, customer, internal business, and innovation and learning perspectives). By providing a "balanced", more integrative view of performance, the new framework promises to provide management with a comprehensive view of the organization allowing it to better engage in strategic thinking and planning. By moving away from strictly financial measures, which emphasize the tangible and financial assets of organizations, the BSC also promises to provide insight into the intangible assets of an organization, such as highly motivated and high-quality personnel, use of best practice processes, excellence in services and high customer satisfaction. The idea is that tracking all these elements potentially provides a better view of the future and long-term best interest of the organization as performance measures are often indicators of future performance. The report from the Accounts Commission for Scotland, Edinburgh, defines the four various perspectives of the BSC as follows:

1) The customer perspective. A public sector organization exists to provide services that meet the needs of its customers (citizens, consumers, clients). It is critical that an organization has clear strategies to meet customer needs and performance measures to assess expectations and levels of satisfaction.

2) The internal business process perspective. To provide quality, and cost-effective, public services an organization must identify key business processes it needs to be good at and then measure its performance in undertaking those processes. This encourages managers to identify key business processes, assess current performance in undertaking those processes, and establish targets for improving performance.

3) The continuous improvement perspective. To achieve continuous improvement in delivering quality, cost-effective services an organization needs to ensure that it is able to learn and to improve from both an individual and organizational perspective. It is important to measure the organization's ability to learn, to cope with change and to improve through its people, its systems and its infrastructure.

4) The financial perspective. Any organization will continue to require key measures of its financial performance but, again, these measures need to be directly linked to the overall goals of the organization.

Figure 2.1: The Balanced Score Card

(Source: The Measures of Success: Developing a Balanced Scorecard to Measure Performance, 1998. Accounts Commission for Scotland, Edinburgh)
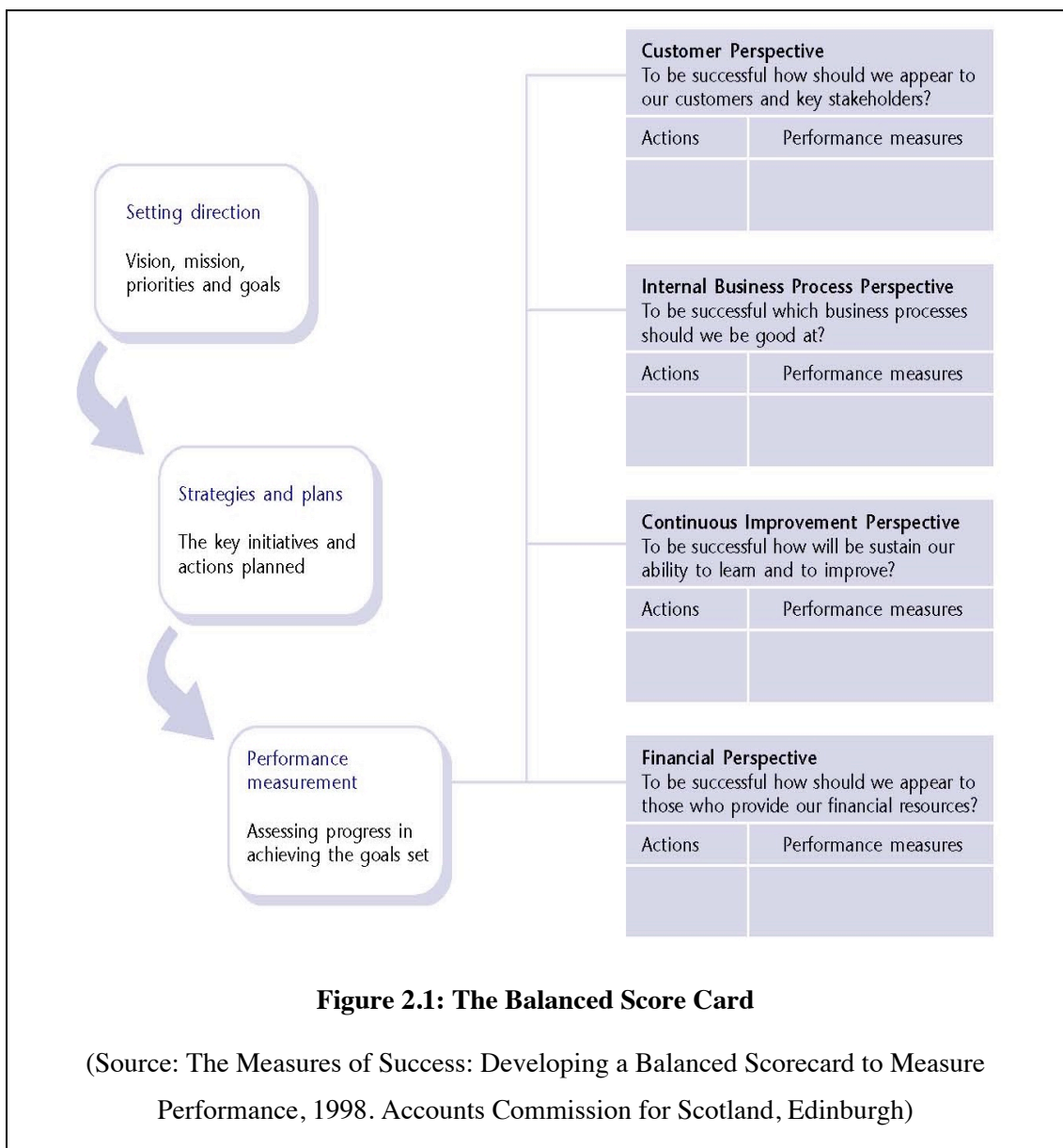
Figure 2.1 above presents a version of the BSC tailored to public organizations. As shown in the figure, the BSC approach starts with an organization's vision, mission, priorities and goals, which are then translated into concrete initiatives that can be tracked and monitored. The process promises to help in establishing and communicating the overall strategic goals of an organization, to identify the key actions and initiatives required to achieve these goals, to asses and measure the impact of initiatives on the organization with a holistic and integrated view through the four perspectives of the BSC, and to monitor progress. As such, the BSC positions itself not only as a performance measurement system framework but also as a strategic management process or approach (Kaplan and Norton 1996).

The BSC seems to be well accepted and quite successful. It is regarded as a performance management framework that contributed to the literature by enlarging the number of interested parties and actors within the process (Gomes and Liddle, 2009). Evidence on the success of the balanced scorecard has been reported across many industries and within the public sector in the US (Hepworth, 1998). According to Kennerley and Neely (2002) findings of several research companies suggests that the BSC has been adopted by a majority of large business: 40 percent of the largest businesses in the USA had adopted the balanced scorecard by the end of 2000 (Gartner research) and 50 percent of surveyed firms worldwide had adopted the balanced scorecard by 2001, with a further 25 percent indicating they were considering it (Balanced Scorecard Collaborative). Since the introduction of the BSC in 1992 the concept has been tested in a multitude of organizations and settings and much further research has been conducted to improve the BSC and/or develop alternative "balanced" performance management frameworks.

However, and despite its apparent successes, the BSC approach faces a number of fundamental criticisms. The feasibility of the BSC as a performance management and measurement framework in the public sector has been questioned by researchers in the field (Gomes and Riddle, 2009). Criticism also focused on the failure of the BSC to include a broad enough view of stakeholders who interact with an organization. Stakeholders such as competitors, end-users, employees, suppliers,

regulators, pressure groups and local communities are not usually accounted for in the BSC although they can have significant impact on an organization (Fitzgerald, 1991; Maisel, 1992; Ewing and Lundhal, 1996; and Lingle and Schiemann, 1996). Much of the discontent seems to stem from the need for performance measurement frameworks that can provide a 'balanced' picture of the business; provide a succinct overview of the organization's performance; be multidimensional; be comprehensive; be integrated both across the organization's functions and through its hierarchy; and explain how results are a function of determinants (Kennerley and Neely, 2002).

2.5.2   The performance prism

Although the BSC was adapted since its inception to better fit the needs of the public sector, it was originally developed with the private sector in mind. The idiosyncrasies of a public organization are most noted in its multiple principals as discussed earlier. Against this background the Performance Prism was proposed by Kennerley and Neely almost a decade after the introduction of the BSC as an alternative performance measurement framework that better meets the needs of the public sector. The views and needs of stakeholders such as competitors, end-users, employees, suppliers, regulators, pressure groups and local communities are accounted for and constitute a key component of this framework. The authors claim that, *"This framework adopts a stakeholder centric view of performance measurement and for this reason is considered to be particularly relevant to the public and non-profit sectors."* (Neely *et al*., 2002)

As illustrated in Figure 2.2 below, the performance prism consists of five distinct but linked perspectives of performance (Neely *et al*., 2001; Neely *et al*., 2002). Namely, these perspectives are: stakeholder satisfaction, strategies, processes, capabilities, and stakeholder contribution. In comparison with the BSC, the performance prism is more inclusive in that it accounts for all stakeholders, not just shareholders and customers. It starts with the question "Who are the important stakeholders in your

organization and what do they want and need?" The second perspective focuses on strategy and asks "What are the strategies we require to ensure the wants and needs of our stakeholders are satisfied?" The authors point out that organizations first have to understand their stakeholders before they can explore the issue of what strategy should be put in place.

Therefore, while the BSC starts with the vision and mission, i.e. the strategy of an organization, the performance prism explicitly begins with the stakeholders. The third perspective, the Processes perspective, asks the question "What are the processes we have to put in place in order to allow our strategies to be delivered?" The fourth perspective relates to capabilities and asks "What capabilities do/shall we require to operate these processes – both now and in the future?" According to Neely, Adams and Crowe (2001), capabilities are the combination of people, practices, technology and infrastructure that together enable execution of the organization's business processes (both now and in the future). The last and fifth perspective focuses on stakeholder contribution and asks the question of what the organization wants from its stakeholders (as opposed to what the stakeholders want from the organization). For example an organization may want loyal employees in return for career development and employment. The authors claim that no other framework recognizes the reciprocal nature of the relationship between stakeholders and the organization.

Figure 2.2: Schematic Representation of the Performance Prism

(Source: Kennerley and Neely, 2001)
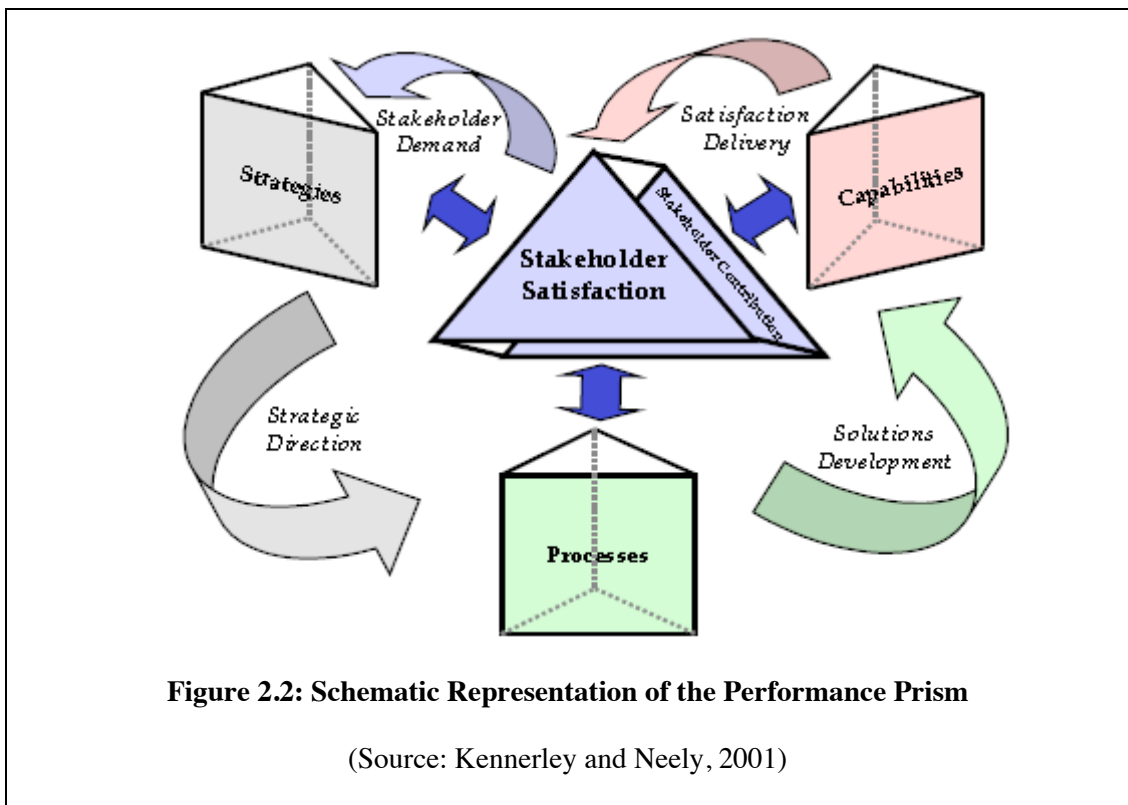
The performance prism is mainly criticized as being too wide a framework. The literature review also reveals that there are very few known implementation of the performance prism rendering it difficult to make any inferences about its pros and cons.

2.5.3 Logic models

The third performance measurement framework is called the Logic modeling. It can be defined a systematic and visual assessment tool that is also commonly used in project and program evaluation. The logical framework was originally developed by the United States Department of Defense and adopted by the United States Agency for International development in the late 1960s (Uribe and Horton, 1993). The framework has since been adapted and widely implemented throughout the world in the private sector (see for example: Rush and Ogborne 1991; Saldannha and Whittle 1998; Kellogg Foundation, 2004) as well as the public sector (McLaughlin and Jordan, 1998; Cooksy *et al.*, 2001). The logic model of an organization or activity

maps what is invested, what is done, to the outcomes. Logic models, in trying to understand input-process-output-outcome relationships, break down the subject (organization, program or initiative) into its constituent components. The models depict assumptions about the resources needed to support activities, and the activities and outputs needed to realize the intended outcomes of a program or initiative. Logic models facilitate thinking, planning and communication about the organization's aims, objectives and accomplishments, making it easier to develop a meaningful evaluation.

The main purpose of a logic model is to provide stakeholders with a roadmap describing the sequence of related events that connect the resources and activities of an organization with the desired outcomes. In this way, logic models serve as a basis from which to analyze or evaluate whether planned actions are likely to lead to the desired or intended results. Logic models are often presented in a similar way to flowcharts.

Figure 2.3: Basic layout of a logic model

(Source: Kellogg Foundation, 2004)
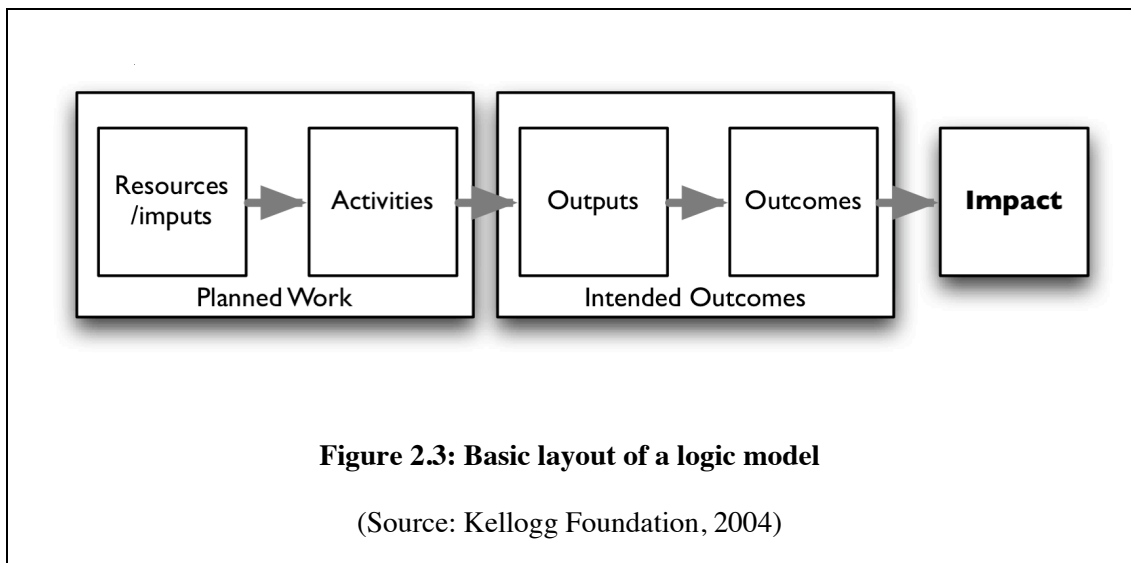
Figure 2.3 highlights the two main subdivisions in the framework structure of logic models:

*1) **Planned work.*** Refers to resources or inputs, be they financial, organizational, or community-based, needed to implement the organization's program(s) and, secondly, to the activities, be they processes, events or tools that are undertaken to produce intended program results.

*2) **Intended outcomes.*** Refer to the desired results of a program or initiative. Outputs and outcomes indicate specific changes associated with activities. Outputs are defined as the direct results of activities, while outcomes refer to desired results or wider intended results, for instance the impact on stakeholders.

Logic models can be applied during all stages an assessment cycle: to explain why the envisioned activities would be successful (designing activities); to map and monitor progress on the stated objectives (performance monitoring and management); and to connect resources and activities to desired outcomes (performance evaluation and accountability).

Logic models have certain advantages over other assessment tools (see for example Cooksy *et al.*, 2001 for a discussion of the advantages). Chief among these advantages is the capacity of logic models to help organize the design, management, and evaluation of an organization or program. They may also provide a structure to allow the evaluator to assess the importance of the context and resources available to the organization or to the initiative, the quality and quantity of implementation, and the effectiveness or magnitude of the outcomes. Furthermore, logic models can serve as a reference point for stakeholders and help to develop a shared understanding between stakeholders of processes employed by the organization or in an initiative.

The flexibility of logic models is also often mentioned as an asset. Logic models can easily be adapted to incorporate different contexts, assumptions, approaches, inputs, resources, processes, and products. This allows stakeholders and evaluators to run through several scenarios to assess what policies are best. Finally, logic models are not static and can be refined as the program develops. Stakeholders and the

evaluators can integrate feedback into the model, as evaluation and the achievement of outcomes influence other program components.

A disadvantage of logic models could be that the one-dimensional linear structure may create a false sense of simplicity of the activities and dynamics of an organization and its environment. Relationships between the components of the model are often complex and interactive. Although it is possible to include such feedback loops in the model, it is difficult to capture this complexity in a logic model. Secondly, logic models can be a very rigid representation of program planning and evaluation and therefore may inhibit flexibility in implementation (Wholey *et al*., 1994). Those involved in evaluation, monitoring and business planning have to guard against using logic models in an overly rigid way or overemphasizing the direct nature of the causal relationship between components of the program or initiative. Finally, an approach using logic models can be costly. Building a logic model takes resources. These costs should be weighed up against the wider benefits of joint stakeholder understanding in addition to its use for evaluation.

2.5.4   The three frameworks in perspective

The discussion above highlights the historical origins, key features, and the advantages and disadvantages of each of the three frameworks. A cross-framework review can help highlight the connection between these frameworks, underscore the differences among them, and lead to some conclusion about the relevance and utility of these frameworks.

The BSC focuses on strategy and is criticized for not including a broad enough view of stakeholders who interact with an organization. On the other hand the Performance Prism focuses less on strategy and is often criticized for having too broad a view of stakeholders. With its high emphasis on strategy, BSC is criticized as primarily designed to provide senior managers with an overall view of

performance; thus, is not intended for, nor applicable, at the factory operations levels (Ghalayini *et al.*, 1997). Although Logical Models follow quite a different paradigm, they can serve as a reference point for stakeholders and help to develop a shared understanding between stakeholders of processes employed by an organization. Because Logical models can easily be adapted to incorporate different contexts, assumptions, inputs, and processes, stakeholders can potentially run through several scenarios to assess what policies are best. Moreover, stakeholders can integrate their feedback into the Logic Model assessment process.

In comparison with the BSC, the Performance Prism is more comprehensive in that it accounts for all stakeholders, not just shareholders and customers. However, since attention is focused on the process of finding the right strategies that performance measurement should be based on, Performance Prism tends to neglect issues such as how the performance measures are going to be realized. This results in giving little concentration to the process of designing the system (Tangen, 2004). On the other hand, the Logical Model follows quite a rigid path in designing the system; too rigid that it runs the risk of inhibiting flexibility in implementation.

Despite the apparent success of BSC (Hepworth, 1998, reports that no failures of the concept were identified), the literature draws attention to the complexity of the BSC, the need for commitment towards accepting it for the success of its application, and the many pitfalls and problems that were identified in practice (Kaplan and Norton, 1996b). On the other hand, the literature offers very few insights as to the implementation of the Logical Model and the relatively new performance prism framework. Performance Prism has been tested in very few cases (e.g. DHL, London Youth, and the House of Fraser), however feedback has been overwhelmingly positive (Neely *et al.*, 2001).

Out of the three frameworks, Logical Models seem to pay the most attention to inputs such as infrastructure. It is also most suited to address issues related to vertical and horizontal integration and cooperation within an organization. Due to its layout, an organization can apply the same highly structured approach within its

different entities allowing for an end of process comparison and identification of linkages. On the other hand, the Performance Prism is especially equipped to ensure quality as it introduces checks at all points of the process.

Finally, an analytic review of the different frameworks suggests that performance measurement frameworks are still evolving and that current frameworks are more or less "work in progress". The shortcomings of the three frameworks become particularly apparent when it comes to implementation related issues. None of the three frameworks seems to be necessarily simple or easy to use; neither do they have a dynamic or proactive approach built into them. Although all three frameworks help focus the selection process, none of them can claim any specific advantage to solving the dilemma of selecting only specific and relevant performance measures; a particularly challenging issue as discussed in the next section. It should also be noted that despite the availability of a number of 'balanced' performance measurement frameworks, few researches have looked into their effectiveness and the economic benefits they yield (Neely *et al*., 2002).

2.6 Challenges Associated with the Implementation and Use of Performance Measurement Systems in the Public Sector

While different performance measurement frameworks were developed and adapted to allegedly respond to the specific needs and requirements of the public sector, a number of problems associated with implementing and using these systems surfaced along the way. The literature includes plenty of reference to the many challenges and obstacles that apparently prohibit performance measurement systems from fulfilling their promise in measuring the outputs and outcomes of government efforts (Smith, 1995; Hatry *et al*., 2003; Van Thiel and Leeuw, 2002; Propper and Wilson, 2003; Wisniewski and Stewart, 2004; Holzer and Kloby, 2005).

A performance measurement system can be regarded as a management tool potentially capable of monitoring and evaluating organizational performance, improving decision-making, fostering fiscal prudence and taking corrective action.

However, translating aspired government performance into measurable results that can be reported internally as well as externally has proven to be highly challenging (Holzer and Kloby, 2005). Selecting which outputs and outcomes of government efforts to measure as well as the most appropriate ways to measure them is proving to be a thorny mission (Neely, 1999). Moreover, and despite the increasing legislative support behind them in a number of countries, the actual use and employment of performance measurement systems is yet a case to be made. In an examination of US counties, Berman and Wang (2000) report that only one third of counties use performance measurement and out of those only one fifth employ high level, advanced methods.

The literature is reacting to the surfacing problems as evidenced in the growing interest in topics related to difficulties and challenges linked to implementing performance measurement systems (Bourne, 2005). Quite a number of challenges and phenomenon are cited in the literature as linked to designing, implementing and using performance measurement systems. The next section organizes these challenges and phenomenon around few major areas: design and framework, quantity of measures, organizational readiness, misuse and other dysfunctional effects. It should be noted that the following discussion does not attempt to distinguish between deliberate and accidental behavior from the organization's end.

2.6.1 Design and framework of performance measurement systems

One challenging area is related to the design of meaningful performance measurement systems and the degree of balance they entail between the different aspects of performance that an organization needs to monitor and track. When a performance measurement system is designed hurriedly and under pressure, and when performance measures are chosen in haste or without sufficient consideration, the end result is a system of poor construct validity that does not measure what it is intended to measure (Davies *et al*., 1999). A mere policy adoption of a performance measurement system by a public organization is not necessarily accompanied by an

internal capacity to develop a theoretical or methodical framework that can guide the design of performance measurement systems.

A related concern is the question of who designs the performance measurement system. Organizations that develop their own systems have more opportunities to manipulate information to their benefit and as a result evoke many of the negative effects associated with performance measurement systems (Van Thiel, 2001). Moreover, the administrative and organizational underpinning of a developed performance measurement system is another important factor. Organizations that heavily use lists of frequently asked questions or tips on how to handle requests from auditors to design their measurement systems increase the possibilities of inducing negative effects (Van Thiel and Leeuw, 2002).

A common practice followed in designing performance measures is based on matching measures up with the goals and objectives of an organization. Stemming mostly from consulting practices, common designs used include the traditional MBO model and the more recent BSC (Johnsen, 2005). These designs commonly follow a path that cascades from the vision and mission of an organizations followed by listing objectives and finding performance measures that can be used to monitor progress towards realizing set objectives. However, there has been some concern regarding the limitations of this method and some researchers regard it as too simplistic to capture the myriad of issues involved (Courty and Marschke, 2003).

Currently, there is an entire subset of the growing literature that focuses on the management processes for designing balanced performance measurement systems; these are management processes that have been developed in the literature, through consultancy experience as well as from action research (Bourne, 2005). Implementing a performance measurement system that is irrelevant or otherwise poorly designed can provide a false sense of security and accomplishments and in the process will misdirect resources and activities (Bouckaert and Peters, 2002). However, this subset of the literature has not yet been matched by research into the

implementation and embedding of the resulting performance measurement systems (Bourne, 2005).

2.6.2 Quantity of performance measures

A relatively related point, this area of concern focuses on the quantity of the selected performance measures compared to the previous area of concern which focuses on the soundness, balance, and quality of the overarching performance measurement system as a whole. The number of performance measures is an important one, as well as whether these measures have been developed for all tasks that have to be carried out by an organization (Van Thiel and Leeuw, 2002). The past two decades have witnessed a proliferation of non-financial indicators that were posed as a remedy to the shortcomings of their strictly accounting-based predecessor.

In addition, public organizations, with their complex nature and multiple stakeholders, are attempting to cover all aspects of performance deemed necessary for performance measurement systems to serve their strategic role. When public organizations endeavor to realize the full spectrum of their strategic role, a difficult and challenging task in and by itself, they often find themselves swamped with performance indicators. A public manager who is keen on embracing a holistic and comprehensive approach can easily sink in the midst of the many hues of non-financial, multidimensional performance measurements.

Measuring for the sake of measurement has the potential of wasting valuable working time, being expensive and financially burdensome, and might lead to frustration with these newly introduced systems. It also results in the negative phenomenon of information overload resulting in the generation of at least some redundant performance reports (Neely, 1999) and rendering performance measurement systems onerous and obstructive in the minds of public managers.

At the same time the literature warns of the negative effects of having few indicators for a limited part of total performance; both excessive and minimalist emphasis on performance measures can result in negative effects (Van Thiel and Leeuw, 2002). The fewer the number of performance measures, the more challenging it becomes to obtain an accurate report of the performance (Meyer and Gupta, 1994). Striking the right balance between too much and too little performance measures seems to be a challenging mission especially in the absence of research in this particular area that would otherwise shed some light onto the subject of quantity.

2.6.3 Readiness of the organization

An organization's readiness plays an important role in facilitating or inhibiting the design and implementation of performance measurement systems (Johnsen, 2005; de Lancer Julnes, 2001). It is important, in discussing the meaning of readiness, to consider the readiness of management as a key topic. Both management capacity and commitment to implementing performance measurement systems are important factors within such discussion. Designing and implementing performance measurement systems require certain set of skills that are not necessarily available in all organizations.

Moreover, data driven planning and decision making requires specific set of skills. Research shows that lack of capacity on how to include performance results in the decision making process adversely affect the utilization of outcome information and prevent otherwise willing public managers from utilizing this information (Hatry *et al.*, 2003). It seems that an era of performance measurement systems plethora characterized as the "audit society" (Power, 2000) was not necessarily accompanied by the kind of capacity building required to bring public managers up to speed on this new practice.

Management commitment can be viewed in both absolute and comparative terms; absolute and total commitment as well a commitment that nonetheless takes into

consideration the priority of implementing performance measures compared to the priority of other projects at hand (Bourne, 2005). Moreover, the literature draws attention to the changes of priorities that can potentially take place in an organization over time. These changes can be initiated by the application of the performance measurement process itself or by other events outside the control of the local management team (Bourne, 2005). As priorities shift, attention paid to the implementation of performance measurement systems by management changes accordingly. Impatience and resistance are also quoted as part of the challenges that can surface during implementation.

A discussion about the readiness of an organization can also touch upon logistical readiness and availability of IT systems to facilitate the collection, tracking and analysis of data (de Lancer Julnes, 2001). Performance measurement calls for considerable data collection, data storage, data analysis, and report production. The absence of adequate IT systems to facilitate such effort can affect an organization's readiness to implement and utilize performance measurement systems.

2.6.4   Misuse of performance measurement systems and other dysfunctional effects

Upon implementation of performance measurement systems, information is produced in the form of statistics, analysis, budgets, reports, strategies and press releases. This 'outcome information' can be used internally and/or externally (Propper and Wilson, 2003) and as a result may be shared with a diverse audience in the public sector to serve a diverse number of reasons and agendas. A notable finding is that performance measurements have an "embarrassment effect" (Mayston, 1985) and can potentially humiliate public managers if reports published, internally or externally, reveal inadequate performance or failure to meet promised goals. In this case it is argued that misuse of outcome information is not only possible but should also be expected.

In addition, different stakeholders can seek information listed in performance reports to use it selectively and strategically in politics. Use of performance information to ends for which it was not intended, e.g., used by senior management or politicians who interpret it in their own ways to suit other purposes, and/or the information is communicated without sufficient explanation or context (Davies *et al*., 1999) is a serious argument that is often used against performance measurement systems.

Appropriately using performance measures to improve decision making in the public sector is likely to be influenced by factors swayed by political and cultural consideration (de Lancer Julnes and Holzer, 2001, Hatry *et al*., 2003). Research shows that public managers who consider using outcome information experience a number of restraints when they do attempt to include results in the decision making process; they either lack authority, may not realize how outcomes data can improve services or may simply choose not to "rock the boat" (Hatry *et al*., 2003).

On the other hand it is quite costly and time consuming to assess the impact of decision-making on the true goals of a public organization. As a result, there seems to be more tolerance towards using imperfect measures in the public sector (Courty and Marschke, 2003). The notion of using imperfect measures denotes that outcome data and statistics can lack relevance. It can be argued that a continuous employment of imperfect performance measures can render them as unworthy of the associated effort and substantiate any tendency not to use their outcome information in the process of decision-making.

The above serves to highlight some of the complex issues surrounding the use of performance outcome information. Smith (1995) has given a list of unintended consequences and gaming behavior that public organizations can partake in to avoid the perceived negative consequences of measuring performance. *Tunnel Vision* (focus on easily quantifiable aspects of performance and leave out aspects of policy implementation that are hard to measure); *Myopia* (emphasis on short term objectives); *Sub-Optimization* (emphasis on the most efficient parts of an organization); *Misrepresentation and Misinterpretation* (hiding ill performance

through either measures) are among the list of unintended behavior identified by Smith (see also Goddard *et al.*, 2000 for a full discussion of how gaming behavior is derived from the principal-agent model that governs the relationship between stakeholders and public organizations).

Gaming behavior captures the notion that the investment allocation that maximizes performance outcomes does not necessarily correspond to the allocation that maximizes value added (Courty and Marschke, 2003). This stems from the view that it is quite difficult to arrive at performance measures that accurately and perfectly capture the goals they are supposed to convey. As a result, an organization ends up with performance measures that do not necessarily convey the true contribution of its employees, or reflect the value added that their work results in bringing about. It follows that these performance measures are likely to elicit dysfunctional responses; employees might misallocate some resources, over-invest in some tasks or under-invest in others.

The situation is further aggravated in organizations that reward certain aspects of performance especially in cases where imperfect measures have been instituted. In essence these organizations are encouraging their employees to focus on the measurable performance and not to spend any effort or time on some activities that can be critical and productive but are not necessarily rewarded (Burgess *et al.*, 2001).

Describing a related but lightly different behavior, LeGrand and Bartlett (1993) tossed the term *Cream Skimming* to depict the behavior of public organizations that discriminate against certain inefficient aspects of public policies by providing services only to those who make the least, or least expensive, use of them. An example of such behavior is excluding chronically ill patients in the health sector (Van Thiel and Leeuw, 2002). Cream Skimming basically gives an inaccurate picture about the success of an organization which defeats the key purposes beyond performance measurement systems.

There are a number of dysfunctional effects, other than those listed above, which face performance measurement systems. There are concerns that performance measures can falsely convey an impression of objective truth, and by making certain aspects of performance visible, they marginalize other management activities (Van Peursem *et al*., 1995). There are also a number of concerns with the use of performance measures in accountability evaluations (Perrin, 2007). Such exercises often lack utility as some governments require evaluations to be carried out as a matter of course with limited purpose. They can have perverse unintended consequences. The costs of evaluations are real and evaluations may involve significant measurement costs, particularly when output is multi-dimensional and complex (Perrin, 2007).

It follows that a significant diversion of resources may result that could have been spent towards improving the organization or services. Performance measurement may also effect culture negatively by putting emphasis on justification and making managers and employees defensive - a learning organization should be focused on improvement which requires an open culture, were managers and staff are constantly questioning what is being done and what can be done better. As a result, the use of performance measures may inhibit innovation and lead to *Ossification*, or organizational paralysis (Smith, 1995). One last point is related to the possible partiality that can take place due to the selection of data sources, respondents and research methods during the implementation of performance measurement systems (Van Thiel and Leeuw, 2002).

2.7 Summary of the Literature Review

There are clearly a variety of concerns associated with performance measurement systems. Unintended and deliberate negative behavior and phenomenon occur for different reasons and at different points of time in a life-cycle of a performance measurement system. Figure 2.4 maps the key challenges described in the discussion above according to the likelihood of their occurrence during the life-cycle

of a performance measurement system. The mapping assumes that the first main stage of introducing a performance measurement system focuses on selecting the general framework within which it will operate and sketching the blueprint of the system. Issues related to design, organizational readiness, ossification, cost, as well as misuse and gaming as discussed earlier in the chapter are likely to be of concern, or even take place, during this initial stage.

The second main stage focuses on the development and selection of specific performance measures that will be used to monitor progress towards set goals and objectives. Issues related to quantity of selected measures, organizational readiness, as well as misuse and gaming are likely to be of most concern during this stage.

The third stage focuses on the actual implementation of the system during which data is collected and analyzed and performance reports are produced. Organizational readiness, cost, as well as misuse and gaming are likely to be the most prominent concerns during this stage. The final stage focuses on utilizing performance outcome information to guide decision and policy making, adjust strategy, reward performance and take corrective measures if necessary. Challenges related to organizational readiness, misuse and gaming are some of the most concerning issues during this stage.
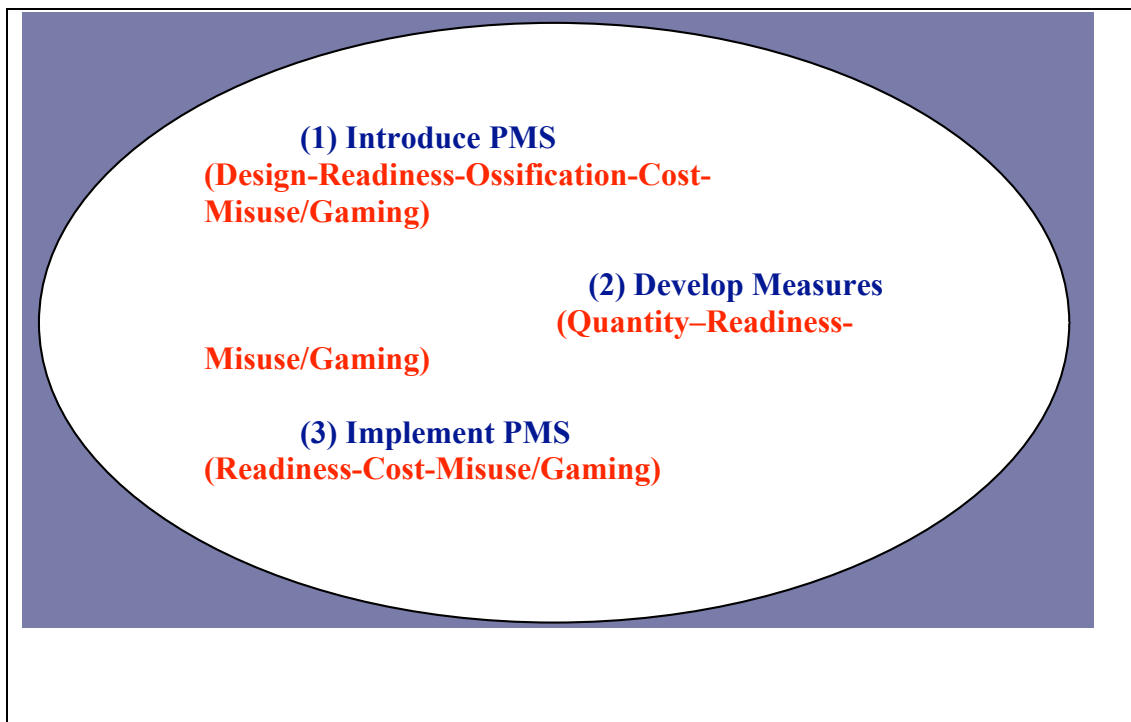
Figure 2.4: Challenges associated with PMS across Different Stages

2.8   The Gap in the Literature

A literature review clearly indicates that much has been written in theory about performance measurement systems; their history and evolution, different approaches, value and effectiveness, and the obstacles to performance measurement. It also reveals a relatively young field that is still in flux; one with characteristics, roles, scopes and boundaries that have not yet taken their final shape.

However, a literature review also shows that less in known in practice about the operation and utilization of performance measurement systems, especially in the public sector (de Lancer Julnes, 2001; Propper and Wilson, 2003). Despite the growing interest in designing balanced performance measurement systems, as evidenced in the growing subset of the literature that focuses on this topic, this work has not yet been matched by research into the implementation and embedding of the resulting performance measurement systems (Bourne, 2005).

In particular, there is little attention to the utilization of measurement data during the stage of implementation: defined in this research as the stage that follows policy adoption of a performance measurement system and the actual design and development of the system. There is little research that examines the patterns of performance measurement use and the factors that influence the utilization of its data in the public sector. The scarcity of research in this area exists despite evidence that externally performance reporting requirements, mandated by governments eager to promote their measurement systems, are not contributing to the improvement of measurement data utilization (de Lancer Julnes and Holzer, 2001). And yet, it is reasonable to suppose that the most significant aspect of a performance measurement system is its impact on implementation and the way in which it is used by stakeholders.

And while a considerable subset of the literature focuses on the different shapes of intended and unintended misuse of outcome information, there is not much research into the actual process of utilization that could otherwise shed much needed light onto these phenomenon and the reasons behind them. Indeed the literature recognizes this gap and warns of the lack of attention to topics related to the utilization of performance measurement systems during the implementation stage (Burgess *et al*., 2002; Bourne *et al*., 2002; Courty and Marschke, 2003). The literature also warns that much of what is identified is only based on anecdotal accounts (Fountain, 1997). Public organizations that seek to benefit from their performance measurement systems need to be concerned not only with designing the systems and arriving at the right mix and quantity of meaningful performance measures but also with the appropriate use and utilization of their outcome information.

Moreover, and despite some attention to the overall importance of understanding the needs of stakeholders as noted mainly in frameworks used to design measurement systems, little attention is paid to exploring and cataloging the actual information needs of stakeholders. For example, the performance prism purposely places the

satisfaction of stakeholders at the heart of designing a performance measurement system. Likewise, the Balanced Scorecard attempts to pay attention to the satisfaction of stakeholders through the customer perspective. However, there is scarcity of research that identifies the end user of measurement information in the public sector (Wisniewski and Stuart, 2004), ascertains the key categories of stakeholders' information expectations and needs, and/or proposes methodical approaches to discover these needs. There is also scarcity of research that examines performance measurement systems from the perspectives of stakeholders in real life situations. Indeed, the literature review clearly indicates that little attention has been devoted to exploring and understanding the performance information needs from the perspective of stakeholders (Wisniewski and Stewart, 2004). Such research can potentially be valuable to those interested in promoting performance measurement systems and improving the utilization of its generated data and information.

CHAPTER 3

METHODOLOGY

3.1   Introduction and Research Questions

A distinction is made between the stages of adopting and implementing a performance measurement system in the public sector and that of utilizing the outcome information of such a system. This research focuses on the information utilization stage. It aims to explore the ways performance measures' outcome information is utilized by key stakeholders and the reasons behind these types of utilizations.

In particular, this research aims to illuminate on the topic of intended and unintended consequences of implementing performance measurement systems by **exploring the key factors that influence and affect utilization of measurement information from the perspective of key stakeholders**. A pertinent case study is examined in depth, and the analysis attempts to explore the model of the utilization process in this specific case and the extant reasons behind the shape and dynamics of this model.

It follows that the overarching research question can be stated as: How was information produced by a performance measurement system used by stakeholders and what were the factors that affected and influenced this utilization process? In an attempt to cover the different phases that might follow the dissemination of outcome information, as well as the different relevant factors that can influence utilization as discussed in the literature, the overarching research question is further broken down into the following research questions:

1) What happened to the data/ information that was collected and produced by the performance measurement system? How was it disseminated to the different stakeholders inside and outside the organization? What was done to share knowledge (emails, informal meetings, formal meetings, briefing, reports, and structured seminars)? Were there any changes that took place in the organization in order to encourage (or discourage) and/or institutionalize a process of utilizing the produced data/ information (training, new processes and revised Manuals of Procedures)?

2) How was the data/information produced by performance measurement systems perceived by the recipients of this information? How was the data interpreted? For example, did it impact important issues from their point of view? Did they see it as presented in a suitable way and in a timely manner that rendered itself useful to guide or influence decision making?

3) How was performance measurement data/information utilized for monitoring, evaluation, planning, reporting, policy, strategic and operational decision making by the key stakeholders and how was it used differently by each of them? How was it used to influence operational decisions (staffing, building …) that reflected the intentions of these stakeholders as well as bring about changes in policies and direction (strategic, sector-related changes), and how was it used by stakeholders in a manner that was not intended by the design of the performance measurement system?

4) Why were the performance measurement data/information utilized that way?

5) What were the factors that affect and influence the process of utilization (challenges, opportunities, facilitating or inhibiting factors)?

As shown in Chapter 2, there is a large volume of research and literature regarding performance management and measurement systems in general and also covering these systems in the public sector. However most of this work focuses on the system

itself, its strengths and weakness, and stops short of covering the stages that follow the collection and dissemination of performance measures data and outcome information. There is little research that examines the implications of the utilization stage and the ways outcome information is used in the public sector. And yet, it is reasonable to suppose that the most significant aspect of a performance measurement system is its impact on implementation and the way in which it is used by stakeholders.

The gap in the literature gets even wider concerning empirical research into the utilization of performance measurement data and information in developing countries. Addressing this gap carries more importance at a time when performance measurement concepts are increasingly gaining ground and being introduced as part of promoting public reform efforts in these countries. There is scant empirical research in this specific area and this research attempts to bridge parts of the existing gap.

Although this research does not replicate or extend a prior study, it does address the highlighted gap in the literature while at the same time builds on prior work by using exploratory constructs from the literature as theoretical propositions to inform the basis of the research questions listed above. Based on the literature review, a prompt list was developed with potential characteristics and factors found to influence the success of performance measurement systems and the utilization of their measurement data. The prompt list was later on incorporated into the questions used to collect primary data as explained in the next chapter.

Despite the many arguments regarding whether the best research is that which proposes knowledge or that which validates knowledge; all research investigations involve a continuous and repetitive cycle of description, explanations, and testing (Meredith, 1989). Any specific research project may involve only one of these stages at a time. It can be argued that exploratory research is required, if not the only way, to achieve descriptive research. For this reason the two terms can be considered closely related and this research will use them interchangeably. The

nature of this research is **exploratory and explanatory** allowing it to cover both the **description and explanation** cycles of research stages. The focus is on **proposing knowledge rather than validating knowledge**; on **exploring and explaining rather than testing theory.** Appropriate research strategy and data collection methods are used to achieve this focus as discussed in the following sections.

3.2   Methodology Paradigms and the Basis for This Research

Traditionally, much research in management has been based on positivist and post positivist paradigms, quantitatively based, along a linear deductive path (Riege, 2003). Viewed as research in the physical sciences, this work is resolutely rooted in a pursuit to uncover aspects of an objective reality and stems from ingrained business school traditions and training programs (Amis *et al*., 2007). In a recent note from the editors of the Academy of Management Journal, Gephart (2004) stated that a large proportion of the **qualitative** research submitted to the Journal that he has reviewed were of positivist or post positivist orientation.

This is interesting in that management and organizational researchers engaged in traditional forms of qualitative research seem to have been trying to align themselves with the traditional positivist and post positivist paradigms. It also means that they have been demonstrating the quality of their work by adopting traditional methods, rooted in a positivistic understanding of the social world. This type of work adopts a position that is termed Foundationalism (Guba and Lincoln, 2005) and is at the center of conversations about the importance of linking considerations of quality in qualitative research with the ontological and epistemological orientations and positions that underpin the scholarship.

At the same time, there is growing evidence of the existence of an ongoing debate about the actual and desired nature and orientation of management research and the role of qualitative research designs used within this research (Pettigrew, 1995; Denzin and Licoln, 2005; Gephart, 2004; Clegg, 2002). There also seems to be some

slow and implicit changes in the position of management research towards the character of the knowledge production paradigm (Gibbons *et al.*, 1994) and the beginning of "assuming a realist perspective" (Hunt, 1990); this perspective is arguably more oriented to the practitioner point of view and use. And while a claim to an existing shift in paradigms can not be made, alternative conceptions of research seem to be surfacing even if suggested epistemologies do not adhere to the dominant norms of the field as it stands today.

Management research is concerned with building a body of knowledge which documents, codifies, and conveys a problem and solution relating to understanding and improving the practice of management. The field's concern with understanding the organization explains the large portion of its research that is based on casework where generalizations are arrived at from certain managerial routines. Using Becher's taxonomy as an analytical tool to dimensionalize and characterize an academic discipline, Tranfield and Starkey (1998) concludes that management research is a soft, divergent, applied and rural field.

The soft dimension of management research is highlighted by the field's salient high tolerance of a wide range of ontological and epistemological paradigms. As a result, the field is characterized by a lack of concurrence on quality standards. This is somewhat linked to, and might be a result of, the divergent feature of the field. The discipline includes a wide range of values and ideologies and its actual boundaries are not clearly defined. This means that any given phenomena can be examined utilizing different methodological methods. The applied dimension of management research means that this is an action-oriented discipline with a focus on studying phenomena concerned with both design and implementation elements. Hence, the followed research process needs to be based on sound empirical grounds. It also needs to employ collaborative connections with practitioners to ensure validity during the data collection and codification stages.

The above highlights the idiosyncrasies of management research and the necessity to strike a careful balance between being theory anchored and ensuring practice-

orientation. The field should have no room for poor-quality empirical research or sloppy theorizing (Huff *et al*., 2006). At the same time, concern with application of findings as well as the issue of user relevance can end up dictating the subject matter on the research agenda. Pettigrew's (1995) notion of dual hurdle of "embeddedness in the social sciences and the worlds of policy and practice" identifies academic rigour and practitioner/policy relevance as major challenges facing management research. Moreover, if management research indeed utilizes research methods from associated disciplines, and if it operates no single, agreed upon, ontological or epistemological paradigm, researchers are often faced with the difficult question as to the most appropriate research process they should follow.

As a starting point, it is important to review a brief description of the different key paradigms to research methodology and highlight the main characteristics of each paradigm. This will help set the selected research method that this research opted to use in perspective to the different available research paradigms.

Four theoretical paradigms to research methodology can be identified as the most significant and widely recognized paradigms. These are: positivism and post-positivism, realism, critical theory, and constructivism.

Positivists take the view that natural and social sciences are composed of a set of specific methods and research should discover and measure independent facts about a single apprehendable reality driven by natural laws and mechanisms. They adopt a position that assumes the existence of a reality that can be uncovered, documented and not contaminated by the researcher. In the world of positivism, the aim of science is to build up objective and causal relationships that demonstrate how constructs of discrete elements work and perform from a relatively secure base, taking quite a broad view of the matter. Postpositivists accept that it is almost impossible to perfectly realize the reality but steps can be taken by the conscientious researcher to minimize the effects of the researcher on the findings generated (Amis *et al*., 2007). The terms positivism and post-positivism are used interchangeably here despite the differences between them. Positivism is typically exemplified by a

deductive method of inquiry seeking for theory confirmation in value-free, statistical generalizations (Riege 2003; Tsoukas, 1989). As shown earlier, an ongoing debate is casting doubt about the prevailing orthodoxy in management research which is based on positivist theoretical paradigm. The very view of anchoring management research in the physical science is increasingly countered. The discipline's diverse nature, its concern with understanding complex organizational attributes and people's understandings of issues, and its key goal of improving the links between theory and practice renders this paradigm unresponsive to the particular characteristics of management research in general and this research in particular.

Realists take the view that natural and social sciences are capable of unveiling reality albeit not with certainty. In other words, realists realize that there are differences between the real world and their particular view of it. They attempt to construct the various views of this reality and aspire to comprehend phenomena in terms of which ones are relative in place and time (Riege, 2003). The main distinctive characteristic that sets realists apart from positivist is that realism does not rely as much on deductive research inquiries. Realism views research methods that are more inductive in nature as more appropriate for exploring, discovering and building theory than testing theory through analytical generalizations. This research is concerned with discovering new relationships of realities, building up an understanding of the meanings of experiences, and transforming people's experiences into verbal experiences of the researcher. In that sense, a realist theoretical paradigm seems to offer a better fit to the specific characteristics of this research as will be discussed further under "Selected Research Strategy" section below.

The critical theory paradigm assumes apprehendable social, political, cultural or economic realities and incorporates a number of historical or virtual structures of these realities that are basically taken as real. Researchers and their investigated subjects are linked interactively, with the belief system of the researcher influencing the inquiry; one that requires a dialogue between the researcher and his/her subject.

This means that objective or value-neutral knowledge does not exist because all claims are relative to the values and beliefs of the researcher (Riege, 2003).

Constructivism, which is somewhat similar to critical theory, makes assumptions that are subjective. However the created knowledge depends on the interaction between and among the researcher and the respondent with the aim of increasing an understanding of the similarities and differences of constructions that both the researcher and the respondent initially held in order to become more aware of and informed about the content and meaning of these constructions (Anderson, 1986). In that sense, the essence of constructivism is multiple apprehendable realities, which are socially and empirically based, and intangible mental constructions of individual persons (Riege, 2003). Constructivists believe that knowledge is theory-driven. They also believe that separation between theory and practice is not possible and neither is a separation of researcher and research object or subject. In short, both methodologies of critical theory and constructivism are dialectical which means that they are focused on an understanding and reconstruction of the beliefs that individual persons initially hold, trying to achieve a consensus by still being open to new interpretations as information and sophistication improve (Guba and Lincoln, 1994).

3.3   Selected Research Strategy

In relation to this research, a strategy can be defined as a way of investigating an empirical topic by following a set of pre-specified procedures (Yin, 2003). If applied with appropriate scruples and commitment, it was claimed, the production of knowledge of the world would be guaranteed (Hughes and Sharrock, 1997). Given that each research strategy utilizes different ways of collecting and analyzing empirical evidence and has its own advantages and disadvantages, the selection of an appropriate strategy is essential to avoid misfits between adopted research strategy and the research goals. This selection should be guided by the **nature of the research questions**, the **requirements for control over behaviour**, and **the focus on temporary events** (Yin, 1994; Benbasat, 1984). The rest of this section explains

how the selection of a research strategy took into account all these guiding points and how the selected strategy fits the orientation, nature, and focus of this research.

This research selects the **case study** as a research strategy utilizing mainly qualitative evidence. The case study method originates in the social sciences, particularly in the fields of sociology and anthropology (GAO, 1990). The case study involves an in-depth examination of an instance. It can be used for "*learning about a complex instance, based on a comprehensive understanding of that instance obtained by extensive description and analysis of that instance taken as a whole and in its context*" (*GAO, 1990: page 15*). A complex instance refers to a situation where the relationship between the input and output is not clear or readily understandable. A number of influencing factors could be affecting this relationship but the exact nature of that effect, and the causal relationships, are ambiguous and not clearly understood.

One of the key benefits of a case study is that it offers the opportunity for a detailed analysis of a real-life setting where context and system character are believed to be very important. This makes it quite fitting for management research, carrying the feature of an applied discipline, which is characterized by a relatively high utilization of casework where generalizations are drawn from restricted managerial practices and realities (Tranfield and Starkey, 1998).

In general, qualitative methods, such as the use of a case study, follow the realism theoretical paradigm and its modes of inquiry since the main objectives are to discover new relationships of realities and build up an understanding of the meanings of experiences rather than verify predetermined hypotheses (Riege 2003; Hunt 1990). A realist perspective is arguably more practitioner-oriented which is more appropriate for the purposes of this research. In addition, the phenomena under investigation, in the context that it exists within, has not been fully discovered and comprehended. Realists' investigation often seems more appropriate to identify phenomena and transform people's experiences into verbal experiences of the researcher (Tsoukas, 1989).

This research attempts to comprehend phenomena in terms of which of the various views of reality are relative in place and time. It aims at presenting an accurate version of the observable facts noted in the case under study, providing some consideration of alternative explanations of these observable facts, and offering a conclusion based on the single explanation that appears most congruent with them. Moreover, there is limited volume of research and literature regarding the specific subject under investigation. Given all the above it seems more appropriate to follow inductive rather than deductive research inquiries in an attempt to explore theory rather than test theory through analytical generalizations. Connections are made from the empirical world to the theoretical world.

As a research strategy, the distinguishing characteristic of the case study is that it attempts to examine a contemporary phenomenon in its real-life context, especially when the boundaries between the phenomenon and context are not clearly evident (Yin, 1981). The selected case is a current, and young, organization that is examined in its real-life context. Context in the selected case study organization is highly relevant for this research and the boundaries between the phenomenon under examination and context are not clearly evident. Although this research assumes that the organizational context, design process and the content of resulting strategy (the performance measurement system), had potentially affected implementation and utilization of measurement data, the exact nature of that effect as well as the causal relationships are ambiguous and not clearly understood.

Moreover, the types of questions posed by this research are a mix of explorations and explanations (how and why) rather than incidence questions; these kinds of questions are typically best addressed by case studies as opposed to other research strategies (Yin, 1981). Also, many researchers believe that the purpose of the case study method is to generate hypotheses rather than test or confirm them (GAO, 1990). Since the focus of this research is on proposing knowledge rather than validating knowledge, a case study method seems to be the most appropriate strategy to adopt.

Another key argument for the use of a case study method as a research strategy is related to the focus of the research on understanding the performance utilization process from the perspectives of different stakeholders. This calls for understanding the different 'realities' of the many stakeholders involved in this complex situation. These stakeholders are likely to have their own perceptions of the utilization process and even their own explanations of the reasons that affect that process. Consequently, it is much more appropriate to use a method that has the capacity to appreciate, characterize, and reconstruct these multiple realities, such as the case study method, rather than one that assumes the existence of a single 'truth' (GAO, 1990). Finally, this research does not intend to control actual behavioural events and its focus is on contemporary events offering further justifications for the use of the case study as a research strategy.

3.4   Limitations of Selected Strategy and Counter Arguments across Two Paradigms

As noted above, there are a number of advantages associated with the use of the case study. The case study allows for in-depth examination and understanding of a contemporary occurrence in its genuine context making it possible to build up a deep understanding of the meaning of real-life experiences. As a research strategy it offers a unique opportunity to address the kind of questions posed in a research interested in exploring and explaining a real-life, contemporary phenomenon. However this method is not without limitations and challenges and most of the criticism it faces seems to focus on two main areas: the generalizability of its findings as well as the validity and reliability of the method itself. Although generalization of any particular research findings is typically addressed as part of external validity, the scope of criticism warrants paying special attention to this point and addressing it as a separate topic in the following discussion.

3.4.1 Generalization

In addressing this common criticism of the case study, it is crucial to revisit the underlying mission that this method is supposed to serve. A case study method is rooted in the need to comprehend a real life phenomenon. The focus of the researcher in this method should be on attaining new in-depth understandings of an organizational phenomenon, identifications of patterns in data, explanations and interpretations about the rich practical experiences of the practitioners; experiences that were until that time mostly unknown and unexplained.

Campbell (1975) argues that search for an explanation is a kind of pattern-matching process. This process can be applied even if there is only a single case because the pattern must fit multiple implications derived from an explanation or theory. Hence it would be *incorrect to judge the particular situation of a case study research by the norms of experimental design. Otherwise this could stipulate that a single case study (or even a small group of cases) could never provide a compelling rationale for establishing the importance of a single observation or variable*. The focus here is on the fact that an explanation, and not a single observation or variable, is what is being tested, and this accounts for the frequent outcome where: *Even in a single qualitative case study the conscientious social scientist often finds no explanation that seems satisfactory. Such an outcome would be impossible if (single factors were being tested) - there would instead be a surfeit of subjectively compelling explanations* (Campbell, 1975: 182).

In short, a single case study should not be mistaken for a single observation. Moreover, and despite the difficulty to generalize results of a case study, this is a qualitative research method intended to provide a detailed account of a particular case situation. Although the method has minimal claims to generalization, neither are its findings intended for generalization or representation of other organizations, it can still highlight and enrich relevant concepts and hence offer normative suggestions without precise prescriptions (Hoskisson *et al.*, 1999).

In addition to the above arguments for case studies in face of criticism related to generalization it can be stated that the selected real-life phenomenon in this research can be seen as significant enough to warrant exploring. The case under investigation represents the entire education sector in a developing country and is part of an ambitious reform endeavour that aims at overhauling that sector and introducing innovative concepts and practices. As such, management decision making in the selected case carries a great weight and has important ramifications that affect the education sector in particular as well as the country in general. This, in and by itself, merits the efforts of this research to examine the role of performance measures outcome information in improving rational decision-making in administrative and political processes.

Furthermore, the education sector is one that currently draws momentous regional and international attention. Attempts to reform this sector are monitored by many policy makers, practitioners, researchers, consultants and governments around the world. Reforming the education sector is also viewed by many as part of the remedy to many of the world's current ailments including fundamentalism and terrorism.

It is also worth mentioning that the selected case shares some predominant characteristics with other reform endeavours in the region such as embracement of top-down approaches and dependence on foreign consultants to design and implement reform. Understanding and explaining the previously unknown and unexplored practitioners' rich experiences can offer valuable insights to other similar reform endeavours in the region. Findings from this research are likely to illuminate on an area of growing interest and offer some explanations that can benefit policy makers and practitioners who face similar circumstances in their own domains and countries.

3.4.2 Validity and reliability

Concepts of validity and reliability, also known as "design tests", are quite well known from quantitative research approaches and are deeply rooted in the positivism paradigm. Qualitative methods, such as case studies, generally follow realistic modes of inquiry and have their own concepts/design tests. In addressing concerns regarding the validity and reliability of this method, the epistemological orientation of this research as well as the uniqueness of the single case study research need to be kept in mind.

Since the focus of the researcher in this method is to attain new understandings and explanations about the rich practical experiences of the practitioners' experience then it can be argued that such understandings and explanations can stem *from creative discovery as much as research design* (Riege, 2003). This means that traditional tests and techniques used to establish validity and reliability are not necessarily the most crucial drivers of rigorous case study research and could even suppress the discovery of new meaningful insights and as a result not maximize the quality of the research (Riege, 2003).

Having stated the above, this does not mean that ample attention should not be given to incorporate design tests intended to establish validity and reliability and improve the quality of the case study method. The validity and reliability of case study research, and the resulting established confidence in the data collected, is a crucial concern for both management research practitioners as well as academics.

More importantly, a higher degree of validity and reliability can be translated into potentially more confidence in the possibility of successfully applying and utilizing the research results during the managerial decision-making process. The literature offers a number of suggestions regarding techniques that can be used to enhance the validity and reliability of qualitative research although there is lesser attention to the case study method and even quite scant advice in terms of suggesting a single,

coherent set of validity and reliability tests for each research stage in case study research (Healy and Perry, 2000; Riege, 2003).

In summary, although considerations of quality in this qualitative research cannot be separated from its epistemological orientation, it remains crucial to identify and use appropriate tests and techniques to establish validity and reliability of obtained data and to determine the quality of the research. The next section is organized around each of the major concepts of validity and reliability as offered in the positivism paradigm and their corresponding concepts in the realism paradigm. For each concept a discussion is presented about what it is concerned with and how it is viewed and applied in case study research. This is followed by listing an inventory of corresponding techniques that can be used to establish and test the concept as suggested in the literature. Techniques that were used in this research are briefly identified at the end of each discussion but covered in more details throughout chapters 4 and 6.

3.4.2.1 Construct validity

The concept of validity in qualitative research is generally linked to the notion of ensuring that arrived at conclusions are defensible, credible, and warranted. It comprises of construct, internal and external validity. Construct validity is concerned with establishing appropriate operational measures for the theoretical concepts and propositions that a research is concerned with examining (Yin, 1994). In other words, in designing a study, the researcher must ensure that all the research questions are fully addressed. According to Yin (1994, p.34), the researcher need to ensure that two key steps are covered: (1) the specific types of changes that are to be examined (in relation to the original objectives of the study) are carefully selected, and, (2) the study can clearly demonstrate that the selected measures of these changes do indeed reflect the specific type of change that have been selected.

This concept is close to the concept of **confirmability** in the realism paradigm (Yin, 1994; Riege, 2003). Confirmability in the literature refers to objectivity and neutrality; to whether the interpretations made of the collected data are drawn in an unprejudiced and logical manner.

One of the major worries regarding the case study method is that researchers usually have direct interactions and personal contact with organizations and people being researched. This makes the method perceived as supposedly more subjective than other kinds of qualitative research. As far as construct validity is concerned, and in order to enhance this type of validity, this calls for continuous and conscientious effort from the researcher end to steer away from making subjective judgments in particular during the research design and data collection stages (Riege, 2003).

Challenges facing establishing construct validity were particularly worrisome in this research since the researcher has worked, and continues to work, in the capacity of an external consultant to the organization that is the case under study. Moreover, the researcher has a working relationship with a number of the subjects who provided the data. The researcher was constantly aware of her own perceptions and prejudices and adopted a number of measures to avoid the potential pitfalls that can follow if personal perceptions were allowed to interfere with interpreting the collected data.

The literature offers a number of techniques to establish construct validity in qualitative research. Miles and Huberman (1994) offer some guiding questions that need to be constantly asked in qualitative research in order to establish construct validity. The questions focus on the explicit and detailed description of the study's general methods and procedures to ensure that it could be replicated by others; providing a complete picture through thick description of each step of the data collection and analysis stages; and retaining study data to be available for reanalysis by other researchers.

To further increase construct validity, the literature suggests other techniques including the use of multiple sources of evidence (data triangulation) for protection

against researcher bias (Flick, 1992), establishing a chain of evidence (Riege, 2003), and asking key informants to review draft reports during the report writing phase (Yin, 1994). Other qualitative research techniques that can be used during data collection and analysis stages include conducting confirmability audit such as examining the data findings, interpretations, and recommendations (Lincoln and Guba, 1985; Riege, 2003). More specifically, the confirmability audit calls for the retention of the raw data such as field and interview notes for later inspection by the auditor if required. This can allow an auditor in a future stage to check the quality of the findings and make a judgment regarding how logical were the inferences made by the researcher (Riege, 2003).

Chapter 2 and 3 demonstrate how this research addressed Yin's (1994) concerns regarding the measures used in the course of the study through clearly defining the overarching research issue about performance measurement, describing the scope and domain of the research problem about measurement data utilization, and finally dismantling the overarching research issue of utilization and the specific problem of measurement data utilization into detailed research questions. This chapter and chapter 4 provide details on how this research used data source triangulation to increase construct validity. Chapter 4 makes reference to the fact that raw data and records of data interviews and focus group are retained by the researcher. Chapter 4 also describes how key informants were asked to review transcripts from their interviews before the researcher started analyzing the data. Chapter 6 offers a full review of all techniques used throughout the research study to increase construct validity.

3.4.2.2 Internal validity

Internal validity is an important concept in quantitative research that is concerned with establishing cause and effect relationships in which the researcher can determine that certain conditions lead to other conditions (Yin, 1994). Internal validity is analogous to the concept of **credibility** in the realism paradigm (Yin,

1994; Riege, 2003). Other terms used to express internal validity include true value and trustworthiness (GAO, 1990). Once again the idiosyncrasies of a case study method call for differentiating between this and other research methods as far as what makes a research process internally valid.

Constructing an internally valid research process in a case study research is not necessarily focused on establishing cause and effect relationships but rather on using a credible way to establish or explain phenomena. A researcher in a case study method is continuously looking for patterns of similarities and differences between subjects' responses, beliefs, perceptions and experiences while at the same time trying to pinpoint those elements and pieces that are important for the patterns under examination and endeavoring to explain the mechanisms that produced them (Riege, 2003). In other words, case study research is concerned with uncovering generative means and being confident that research conclusions about a real-life phenomena are arrived at in a credible way. This is quite different than establishing cause and effect relationships since the concept of establishing internal validity in the case study is really related to establishing or explaining phenomena in a rather credible manner.

Some of the tests and techniques suggested in the literature for establishing internal validity in qualitative research include conducting explanation-building and pattern-matching (Yin, 1994) to achieve internal coherence of findings and ensure that concepts are systematically related. Moreover, conducting within-case analysis followed by cross-case pattern matching is suggested to increase internal validity. Other qualitative techniques include peer debriefing, triangulation of sources and methods (Lincoln and Guba, 1985), and researcher self-monitoring during data collection and data analysis (Merriam, 1988; Riege, 2003).

This chapter and chapter 4 illustrate the use of data source triangulation during the data collection stage. Chapter 4 describes the use of flip charts and the coding sheets to detect patterns and build explanation for uncovered patterns and the mechanism that might have produced them during the data analysis stage. There is reference to the researcher's self-monitoring technique throughout this thesis by demonstrating

how adherence to the guidance of the literature was carried out during data collection and analysis.

3.4.2.3 External validity

This is another concept that is well known in the realms of quantitative research. The concept of external validity is comparable to the concept of **transferability** in the realism paradigm (Riege, 2003). Other terms used to express external validity include applicability and generalizability (GAO, 1990). External validity is mainly concerned with the possibility of extrapolating the findings of any given research findings beyond the immediate form of inquiry and into some use beyond the specific. Yin (1994) argues that external validity refers to the domain to which the findings of the research can be generalized.

As discussed in section 3.4.1 about generalizability of findings, the case study method is a qualitative research method intended to provide a detailed account of a particular case situation. It is not concerned with producing universal knowledge or representing other organizations. Consequently, establishing external validity in case study research is quite distinct from the typical statistical generalization that quantitative research might focus on to establish this function.

Even though the findings of a case study could be generalized to some broader theory, the focus remains on analytical generalization versus statistical generalization. The researcher in quantitative studies can utilize the *theoretical replication logic* to verify if patterns and concepts from one particular case can be applied to other settings (Eisenhardt, 1989a; Yin, 1994; Meredith *et al.*, 1998). It follows that a clear definition of the scope and boundaries in the research design phase is called for in order for the researcher to achieve a reasonable analytical generalization (Marshall and Rossman, 1989).

The literature offers a number of techniques for establishing external validity in the case study method. These include comparison of evidence with the extant literature during the data analysis stage in order to highlight the contributions of the research and generalize those within the scope and boundaries of the research, versus to a larger population (Yin, 1994). The literature also recommends the use of predetermined questions, the development of a case study data base, and the use of thick description to describe the research process (Lincoln and Guba, 1985). Moreover, the use of specific procedures for coding and analysis such as symbols and signs is considered a technique that helps achieve external validity (Yin, 1994).

Chapter 3 highlights the use of predetermined questions during the research design stage. Chapters 3 and 4 offer a thick description of the data collection stage, and the specific procedures that were utilized for coding and analysis during the data analysis stage. The discussion in Chapter 4 reflects the continuous effort of the researcher to compare between the set of initially identified theoretical constructs and the empirical results of the study during the data analysis stage in order to better understand and explore the constructs. The conclusion section in Chapter 6 makes reference to the possible generalization of findings to other environments, within the boundaries of the research. Even though the initial focus of this study was on a particular case of education sector reform, the final findings regarding the factors that influence utilization are not specific to the case study and can lend themselves to similar reform experiences that are concerned with the utilization of performance measurement data.

3.4.2.4 Reliability

This concept is concerned with demonstrating that the procedures and processes of the research can be repeated by other researchers and can then achieve similar results. It is related and analogous to the concept of **dependability** in the theoretical paradigm of realism (Yin, 1994; Riege, 2003). Other terms used to express reliability include consistency, explainable instability, and replicability (GAO,

1990). In essence, this concept is concerned with the transparency of the process and conclusions that the researcher followed to collect data as well as make sense out of this collected data. The end goal is that other researchers are likely to achieve similar results, and draw similar conclusions, if they follow the same operations and procedures of the research (Miles and Huberman, 1994; Yin, 1994).

The case study method is a qualitative research that deals with people in real life situations. It is hard to assume that people are static and will provide the exact answers, for example, even if the researcher made sure that others can follow each process and procedure taken in the research. On the other hand, these potentially different results can at the same time enrich the findings about the case being researched by providing additional information and covering different points that were not captured in the original research. The key to achieving reliability in qualitative research is to demonstrate that the process of data generation and documentation was thorough, careful, accurate, as well as appropriate to answer the research questions (Yin, 1994; Riege, 2003).

In order to achieve this, one technique is to give full account of theories and ideas and assure congruence between research issues and features of the study design (Yin, 1994). Giving full account of the theories and ideas for each research phase is also recommended to increase reliability (Le Compte and Goetz, 1982). A careful development and refinement of case study protocol during the research design stage by conducting pilot studies is yet another recommended technique (Eisenhardt, 1989a; Yin, 1994). A researcher can also seek the counsel of other researchers, ensure that observations and actions are recorded as concretely as possible, use the developed case study protocol, assure meaningful parallelism of findings across multiple data sources, and in general ensure keeping a good record of data which includes mechanically developing case study data base (Le Compte and Goetz, 1982; Eisenhardt, 1989; Yin, 1994).

Moreover the researcher can benefit from peer review and examination during the data analysis stage. Giving full account of theories and ideas is also recommended

as one of the techniques to establish reliability from the research design through the data analysis stages. Other suggested qualitative methods techniques include conducting dependability audits to examine and document the process of inquiry as well as clarifying the researcher's theoretical position and biases; both of which can be used during the research design stage (Lincoln and Guba, 1985; Hirschman, 1986).

In order to allow future investigators to repeat this study, Chapter 4 offers a careful documentation of the processes and procedures that were used during the data collection and analysis stages. A description of the case study database including archival records and collected data Excel Sheets is also provided in chapter 4. The interview protocol is included in Appendix 2. The research also identifies and clarifies potential biases by disclosing the working relationship that existed between the researcher and some of the key stakeholders. Safeguarding against the researcher's theoretical position and bias is achieved through vigilant awareness of quality tests and adherence to guidance in the literature regarding the different quality criteria.

3.5   Data Collection

Data in this research came partially from archival records including initial reform publications, white papers, reports, and newspapers articles in order to build a historical account of the case under study. This collection of background data on the original design of the reform, the organizational context, the resulting content of the strategy and the developed performance measures provided an anchor point and served to explain the context of this case study. It also allowed for the collection and use of relevant secondary data. However, **most of the evidence** for this research came from primary data that was gathered using personal interviews and focus groups as data collection instruments.

The use of multiple respondents and multiple resources of data in a case study method as well as triangulation by multiple methods and types of data allows for greater substantiation of constructs and hypotheses (Eisenhardt, 1989a). Respondents in this research represented a varying range of stakeholders that included a policy setter, senior directors, senior management teams, consultants, teachers and school's administrators, as well as members of the media. A full discussion of the selection of the stakeholders is covered in Chapter 4. A mix of in-depth personal interviews and a focus group was used to collect information from respondents who represented most of the identified stakeholders groups.

Triangulation aims at compensating for the weaknesses inherent in each of the used method and dataset by the strengths of the other method and dataset used. In this research, one concern has been that at least some, if not most, respondents might not be willing to report accurately on a number of the issues addressed in the interview questions. The highly politicized setting of the case under study, and possible perceptions regarding risks associated with divulging information, were believed to cause respondents to be hesitant, or perhaps unwilling, to provide faithful information. The overall accuracy of the generated data would hence be influenced by this hesitation to provide accurate reports (Miller *et al*., 1997). The use of multiple respondents and a mix of data collection methods seek to address this concern and allow for a better substantiation of constructs and hypotheses.

3.5.1 Using personal interviews and focus group as data collection instruments

In-depth personal interviews that lasted for about three hours on average were used in order to collect information on many aspects of the phenomenon. These allowed for delving into the questions at hand and were particularly appropriate and applicable to the case study organization were there was a complex subject matter, detailed information was sought, respondents were busy/high status, and there was a highly sensitive subject matter (Berkowitz, 1996). The topic of performance

measurement systems in the public sector is typically highly publicized and involves multiple stakeholders who are likely to have different perspectives and ideas on the matter. Using in-depth personal interviews aimed to document the different perspectives of different stakeholders groups.

Despite its many advantages, personal interviewing was not viewed as the best method to collect data from all of the identified key stakeholders groups. A number of interviewees represented, as a group rather than as individuals, a key stakeholder. The collective perspective of the teachers and schools administrators group, for example, was believed to be more relevant to the questions posed in this research than the individual perspectives of its members. In these cases, the use of focus group as a data collection method was thought to be a more appropriate method to provide another level of data gathering or perspective on the research questions that could not be provided through individual interviews. Furthermore, the use of the focus group to collect data supplied this research with additional rich data and complemented the advantages provided by the use of personal interviews. It also aided respondents' recall of events, proved to be stimulating to respondents, and proved to be quite an elaborative process. All these advantages are all well documented in the literature (Denzin and Lincoln, 1994).

In both personal and focus group interviewing, the researcher was aware of the important characteristics of a good interviewer. Mainly, the researcher was aware of the importance of being flexible, objective, empathic, persuasive, and a good listener during the interviews.

Moreover, the researcher was aware of a number of distinctive challenges that are unique to a focus group setting and, as a result, require more refined interviewing skills. In a focus group setting, an interviewer/moderator is often required to direct the interaction and inquiry and conduct systematic questioning of several individuals simultaneously during a single setting. Emerging group culture may prevail over individual opinion and perspective making "group think" a realistic threat to the strengths and positive characteristics of this data collection method. It is also quite

likely that one person, or a small coalition of persons, dominate the group and the discussion hence affecting the quality of the collected data. The interviewer needs to have the required skills to prevent this from happening, encourage disorderly interviewees to participate and make sure that the entire group is responding and participating in the interview (Denzin and Lincoln, 1994). The researcher's working experience in conducting focus groups proved to be quite helpful in navigating through many of these challenges as indicated in the next chapter under section 4.2.4.

3.5.2   Unstructured versus semi-structured interviews

The initial thinking towards conducting interviews was to use open-ended questions in a rather unstructured approach which would help provide this research with as much unconstrained responses as possible in order to explore the how, why and what concerning the utilization process. However there was a concern that using open-ended questions might lead interviewees to focus on a certain aspect and forgo other aspects that this research was interested in addressing. Another concern was that interviewees might regard the interview process as an opportunity to vent and blame others for some of the challenges they face in their young organizations instead of focusing on the issues under investigation. Yet another concern was related to the importance of identifying the similarities and differences in the experiences of different stakeholders. This required that all respondents provide their account on at least a number of standardized questions.

Given all these concerns, the idea of using open-ended questions as part of an unstructured approach was revisited by the researcher. As an alternative, the use of a semi-structured approach was seen as more appropriate to adopt in this study. A semi-structured approach would bring about the advantages of providing strong similarities in the questions asked across interviewees while allowing them enough room to describe their own perspectives and experiences on the subjects being questioned.

3.5.3 Ensuring rigour

The use of interviewing as a data collection method provided detailed and in-depth access to different views and opinions as well as comparative information on complex issues. However, interviewing represented a more personal form of research that carried a number of pitfalls and concerns that this research needed to rigorously watch for and avoid during the process of preparing for and conducting interviews. The importance of properly contemplating "who to ask" and "what to ask", ensuring clarity of questions, steering away from misleading questions or influencing the interviewee were highlighted and considered early on during the course of this research study. In thinking through the issue of "who to ask", it was necessary to properly identify the key stakeholders that might be expected to affect the case study organization. The Power/Interest grid technique was used in order to guide this research in identifying "who" to ask (Ackermann and Eden, 2003). A full discussion of this process is provided in Chapter 4.

In thinking through "what to ask", every effort was made to ensure that the interview questions addressed, and were linked to, the research questions. All questions, sub-questions, and the corresponding probes were derived from the theoretical propositions and other underpinning theory that was relevant to this research. The questions were carefully crafted to ensure the inclusion of all the themes that would used as priori codes later on during the data analysis stage. Moreover, special attention was made to avoid misleading questions and to ensure that all questions were drafted in a clear manner.

After developing the initial list of questions, the researcher asked two researchers/colleagues to review the questions and comment on their clarity and whether or not any of the questions appeared to be misleading in their judgment. A number of comments were made and the initial sets of questions were revised accordingly. This was followed by sending the revised questions to the advisors of this research to get their feedback on the questions.

Upon getting the approval of the advisors, the list of interview questions became ready to be used in a pilot study. The pilot study included three interviewees and the language used in the questions was slightly changed after conducting these interviews based on received feedback. The final list included eighteen, mostly open-ended, interview questions (Appendix 2). Twelve of these interview questions included a number of probes and sub questions to be used after receiving an interviewee's initial response to the question. The goal was to provide reference to the main themes, ensure that the interviewee understand the points being made, gather specific examples and get further insights.

In determining the manner through which responses should be recorded, it was the researcher's judgment that it would be better to steer away from using recording machines or other devices during the course of the interview. The use of such machines was not a common practice, was generally viewed as culturally insensitive, and could possibly be perceived by interviewees as threatening. In order to avoid a possible risk of information withholding by the respondents, hand written notes were used to record the information. Ample space was provided underneath each question on the interview form in order for the researcher to record the interviewee's responses. Data was collected during the interviews through making detailed linear notes as the interview was in progress.

Issues related to privacy and confidentiality were another major concern taking into consideration a number of cultural factors that surround data collection from local people in general as well as the highly politicized setting of the case under study in particular. In addition, the researcher's affiliation with the consulting company that played a major role in designing the reform posed a serious concern about privacy and confidentiality, the interviewees' willingness to divulge information, as well as the quality and accuracy of their responses. Hence, it was important to try to preempt the occurrence of any of these concerns by disclosing the exact goals of the research and how shared information will be used.

An oral consent was developed to introduce the research topic to the interviewees in an objective and factual manner. The oral consent included an introduction of the researcher in her academic capacity and a description of how collected data will be used strictly for purely academic purposes unrelated to the professional job of the researcher. The oral consent promised giving the interviewees the chance to validate their answers and changing the transcripts accordingly. It also promised keeping the identity of the interviewees anonymous to the extent possible. At the end, interviewees would be asked if they have any questions or concerns prior to proceeding with the interview.

Ensuring rigour also calls for the careful development of a protocol to guide and standardize the steps involved in all interviews and focus group to the extent possible. The protocol covered the main actions that the researcher needs to take in preparation for, during, and after conducting these meetings. It can also be used as a checklist for all the documents that the researcher needs to take to the meeting. In addition, the protocol incorporated a measure to establish construct validity through asking key respondents to the personal interviews to review transcripts of their interviews before the researcher started to conduct the data analysis. It covered in ample details how that process of validation will take place. A full copy of the protocol is included in Appendix 2.

The next chapter offers a detailed account of the data collection and data analysis processes.

CHAPTER 4

DATA COLLECTION AND ANALYSIS

4.1   The Secondary Data Collection Process

Data in this research came partially from archival records including initial reform reports, publications, and media records in order to build a historical account of the case under study. Most of the secondary data collection effort, except for the collection of media records, started in December 2006 and continued until August 2007. More archival records were collected and reviewed as they became available during the research study. The result of the initial secondary data collection effort was the compilation of enough background information that allowed the researcher to describe the asserted original design of the reform, the organizational context, the developed performance measurement system, and the early reaction of the community towards the reform strategy.

The compilation of the media records started in December 2006 and continued to take place until the end of the personal interviews data collection effort. The earliest media records that were examined date back to the start of the academic year during which the reform schools became operational (September 2004). At the same time, the researcher wanted to ensure the availability of media records that covered the same time period during which primary data collection was taking place. The objective was to facilitate future cross checking and comparison of data that originated from different sources. The last media records that were included in the analysis were for May 2008.

All paper documents were arranged in well labelled files and in a way that facilitates easy retrieval when needed. Electronic documents were saved on the researcher's

laptop and arranged according to content in a way that allows for easy access to data. For example, the researcher collected and examined 1009 newspapers' articles, editorials, and radio program records for the period between 2004 and 2008 (excluding 2006) that include reference to the reform or the case study organization. These records were separately catalogued by year to facilitate future analysis of trends or themes in media's position towards the reform. The collected secondary data provided an anchor point that allowed the researcher to better understand the original intentions, background and history of key systems relevant to the case study. It also allowed for future verification of, and comparison with, collected primary data. A full bibliography of archival data records is listed in Appendix 1.

The following sections offer an overview of the key findings from the secondary data collection process.

4.1.1 The case study: history and background information

The case study in this research is a publicly-financed, policy-making, public organization that was established in November 2002 to spearhead an ambitious, system-changing, outcome-based, education reform initiative. The mandate of the case study organization states its responsibility over setting and implementing the country's broad policies and short and long term goals for the new reform system. The country granted the case study organization several legal, budgetary, and structural exemptions to help it carry out its new, and rather unorthodox, mission.

The original ideas, principles, and design of the education reform were conceived by the country's top two leaders, with the help of a primary consultant group, and passed on through the initiation of a reform characterized as a top-down initiative. In particular, two of the reform's four guiding tenets, schools' accountability and parental choice, are supported through the establishment of a performance measurement system that can provide the required data.

From an organizational structure point of view, the case study organization has a **Governing Board** in charge of policy setting for the education sector. The Board should take objective, high-level, policy and strategic decisions that are guided by data. One of the key sources of this data is measurement outcome information produced by the case study organization's performance measurement system.

The other relevant structure at the case study organization is the **Operational Entity**. This Entity oversees and supports a new genre of schools, called reform schools hereafter, and ensures they have all required resources for the successful education of students. The first cohort of the reform schools (12 schools) opened in the 2004/2005 academic year (Brewer et al, 2007).

The last relevant key structure at the case study organization is the **Assessment Entity**. This Entity is charged with assessing and evaluating the performance of the reform schools' system to facilitate the continuous development and offering of quality education options in the country. This entity develops and conducts data collection through testing students, monitoring their learning, and evaluating school performance.

The case study organization can best be described as a semi-autonomous organization that operates in a quasi-market environment marked by competitions between schools. Input management is replaced by a results-based orientation as contracts are drawn with schools articulating which key 'tasks' has to be carried out. The schools' performance is expressed in terms of performance indicators. These indicators are supposed to enable the case study organization to measure and evaluate the performance of the schools and also to increase the opportunity to account for performance.

Operating in the midst of a unique set of local, regional, and international circumstances, it is important to review the context within which the case study organization operates in to better understand the opportunities and challenges it faces. Education, as a sector, is a particularly sensitive and highly politicized sector

in the region that the case study organization exists in. International reports about human capital development in the region continue to focus on the significant role of education in shaping the economy and creating political stability in these countries. The 2003 Arab Human Development Report cites lack of Knowledge capital as the main long-term problem faced by countries the region and calls declining quality the most important challenge faced by education in the region (United Nations Development Programme, 2003). The report warns against reliance on material products or national resources noting that the knowledge gap, not the incoming gap, "determines the prospects of countries in today's world economy" (UNDP, 2003, p.35).

Similarly, the World Bank's 2008 report on education reform in the Middle East, *The Road Not Traveled: Education Reform in the Middle East and Africa*, notes that Middle Eastern countries have not been able to "capitalize on the progress made in increasing the level of human capital in the labor force over time" (World Bank, 2008, p.296).

A number of other international reports highlight the importance of education and learning as a key factor to sustain efforts towards human development (*The Dakar Framework for Action*, UNESCO, 2000, P.45), contribute to increasing earning potential and social mobility and improving personal and family health and nutrition (*From Schooling Access to Learning Outcomes An Unfinished Agenda,* World Bank, 2006, p. 3), and buttress democratic institutions and civic engagement (*Expanding Opportunities and Building Competencies for Young People A New Agenda for Secondary Education,* World Bank, 2005, p. 17).

In addition to international attention to the sector, a number of highly critical and emotionally charged articles have been published in the local and regional media since the inception of the reform. The nature of these articles assert to existence of widespread concerns, in the country as well as in the region, about the commencement of a new era of "colonization" disguised in the form of education reform (see section 4.1.3 for more details). General themes about manipulation of

the tender minds of children by new methods of learning and foreign ideologies abound in these articles. In the middle of fears of westernization, loss of national identity and dilution of religious beliefs, many groups in the society are particularly apprehensive about any changes or reforms in the education sector.

Consequently, it is safe to assume that the goals of the reform, as adopted by the case study organization, are not necessarily agreed upon or even widely accepted in that country. More related to this research is the role that this contextual public attitude plays in affecting the dynamics and the ways different key components of the reform system, such as the performance measurement system, function. The unique external context that this case study operates within conceivably have some bearings on the ways performance measurement data are used, partially used or not used at all. This discussion is also important in thinking through the initial list of key stakeholders as it outlines the magnitude that this reform has in the community and helps draw the initial picture of the universe that defines the key stakeholders.

4.1.2 The performance measurement system

The performance measurement system of the case study organization was designed mainly by the primary consulting group that helped design the reform initiative. A Performance measurement system can be carried out at a range of levels at public organizations depending on the purpose behind establishing these systems (Propper and Wilson, 2003). A performance measurement system can, for example, be used to improve the performance of an entire public organization or it can be used to improve the performance of certain units that are under the jurisdiction of the parent organization (Propper and Wilson, 2003).

In the case study organization, the performance measurement system follows the latter model of operation where data is collected for the purposes of monitoring and improving the performance of reform schools, or organizational units, that fall under the jurisdiction of the case study. The system focuses on monitoring, assessing and

improving the performance of the reform schools. The underlying rationale behind the establishment of the measurement system is to improve accountability by drawing public attention to failing schools and linking their poor behavior to a number of punitive actions (which include revoking a school's license to operate). The system also aims at promoting simulated market competition as parents are expected to enroll their children in schools that demonstrate better performance measures. The notion of 'parental choice' is one of the reform's key principles. Since funding is calculated on a per-student basis, it is expected that schools will be motivated to improve their performance as they compete for students.

Initially, the performance measurement system was linked to an explicit incentive scheme that was introduced at the school management level and incorporated into an accountability system, with options for the case study organization to revoke its contractual agreements with schools identified as failing. Under the incentive scheme, direct financial reward would be given to the individual or sub-group that is 'successfully' managing the school.

During the second cohort of the reform schools, this pay-for-performance approach was brought to an end following a change in policy. There is some evidence from the media review and the interviews that this change of policy came as a result of pressure from the media. However, even in the absence of an explicit incentive scheme, it is still the hope of policymakers that the power of sheer publication of measurement data will have an effect on behavior. Schools will continue to be motivated to improve their performance to (1) avoid embarrassment and shame associated with publicly announcing failing schools, and (2) attract more students, and as a result, more funds. The performance measurement system deliberately exerts competitive demands on schools to improve their performance.

The Assessment Entity was charged with managing all aspects of the measurement system including data collection, storage, analysis, and dissemination of findings. A review of the measurement system reveals a wide scope of measures that cover quite a large number of categories. Data collection started immediately after the inception

of the reform and comprised mainly of two key components: an Inclusive School Appraisal Scheme (ISAS), and an Inclusive Educational Appraisal Scheme (IEAS). Table 4.1 outlines the two key components of the measurement system and provides information about the outcome, frequency, and date of each component. Results of ISAS and IEAS are reported to the society through two means of publications: the local media and special reports called School Articles (SA).

The Inclusive Educational Appraisal Scheme (IEAS) is a modern, end of year, standardized testing system that examines the academic results of students. The annual assessment scheme focuses on outcome measures for four major subjects: the National language, a Foreign language, Mathematics, and Science. Students at the reform schools are subjected to this measurement scheme to monitor their achievement and progress in the four subjects. Table 4.1 provides some examples of the kind of output-oriented performance measures collected by IEAS. The system has a significant capacity and potential for measuring student progress and school quality. It also provides important longitudinal information on individual students. The Assessment Entity has been using this testing system since the early 2004 (Brewer et al, 2007). IEAS is a core component of the School Articles (described below) and is also considered in deliberations concerning schools classification and assessment.

Student Academic Achievement
Average native language scale scores (reading, writing, listening and overall)
Percentages of students at each standard level in overall native language
Average foreign language scale scores (reading, writing, listening and overall)
Percentages of students at each standard level in overall foreign language
Average mathematics (overall) and science (overall) scale scores
Percentages of students at each standard level in overall mathematics
Percentages of students at each standard level in overall science

Table 4.1: Examples of measurement data collected by the IEAS

(Source: Report issued by the Assessment Entity, 2009)

The Inclusive School Appraisal Scheme (ISAS) is an annual and comprehensive survey of all reform schools based on data collected by the Assessment Entity throughout the year. The data is collected using Computer Assisted Self Interview, Paper and Pencil Interview, and Enumeration. The survey covers a wide range of domains using input from various stakeholders such as students, parents, teachers, and principals. This survey complements the Inclusive Educational Appraisal Scheme' score results with contextual information on schools. By including a mix of processes and input measures, the survey goes beyond measuring concrete outcomes and helps situate achievement within a larger context. Table 4.2 provides some examples of the kind of process and input-oriented measures collected by ISAS. The Assessment Entity has been conducting this survey since early 2004 (Brewer et al, 2007).

Measures to Monitor Inputs and processes
Teachers' years of teaching experience
Teachers' satisfaction with the quality of professional development
Teacher Turnover indicators
Principals' satisfaction with teachers' subject preparation
Principals' satisfaction with the instructional quality of school's teachers
Parent involvement in school committees
Schools' provisions to enhance its teachers' skills [by school type]
Ratios of students to school facilities
Ratio of total students to total teachers
School extracurricular activities provided [by school type, by school stage]
Average class sizes for selected subjects
Average school days in the school year [by school type]

Table 4.2: Examples of measurement data collected by the ISAS

(Source: Report issued by the Assessment Entity, 2009)

As mentioned earlier, data collected through the above two major components of the performance measurement system are shared with the media upon its release by the Assessment Entity in the case study organization. It is also reported in the **School Articles (SA).** SAs provide an overview of the results of both the Inclusive Educational Appraisal and Inclusive School appraisal Schemes in a user-friendly, standardized format that parents, schools and communities can easily understand and use. The purpose behind publicizing summary statistics of performance measures, SAs, is to provide information to the society about a school's performance allowing public scrutiny and informing parental choice. Another purpose is to improve schools' performance by creating the incentive to improve their students' educational attainment.

System Component	System Outcome	Data Collection Frequency	Data Collection Date
Inclusive Educational Appraisal Scheme (IEAS)	National Education Data	Once a year	Spring
Inclusive School Appraisal Scheme (ISAS)	National Education Data	Once a year	Fall

Table 4.3: Components of the Performance Measurement System

(Source: Documents from the Assessment Entity, 2007)

It is worth noting that the role and functions of the performance measurement system in the case study are very similar to practices followed in both the UK and the USA. The education sector in the UK is also controlled by a relatively high level of public scrutiny, mainly through Ofsted reports and the publication of the league tables, ever since the implementation of the 1988 Education Reform Act (Propper and Wilson, 2003). The 'No Child Left Behind' (NCLB) Act of 2002 put the use of

accountability systems that are based on student test performance into law in the USA (Cullen and Reback, 2002).

Summary of the performance measurement system

The main purpose behind the design of the performance measurement system in the case organization was hence to create a mechanism that can produce information to support and monitor a results-oriented reform. The system's design was purposefully oriented more toward schools outcome versus schools input and process. Schools outcome measures, such as academic results, were emphasized while schools input and process measures, such as quality of teachers and teaching practices observed inside a classroom represented a smaller and less important portion of the measurement data. Outcome measures supposedly capture movement towards the goal of improving student performance; which is the ultimate goal of the education reform according to the designers of the reform.

This focus constituted a significant departure from a prior education system that focused on rote learning and day-to-day teaching practices. Schools and their management teams were offered sizable autonomy. In return, accountability became almost entirely dependant on the results these schools, or their students, achieved as measured by the key components of the performance measurement system. If successful, the performance measurement system can empower parents and communities with information about school performance to facilitate school choice.

Consequently, the performance measurement system can be regarded as a critical component of the reform. This warrants considerable attention to how well this system is functioning and meeting the expectations of the reform. A review of secondary data revealed that the feedback and information that is provided by the collected measurement data is in essence the only means of ensuring that progress is taking place, standards are adhered to, and the principles of reform are being realized.

4.1.3   The role of the media: reactions to performance data

Ever since the country announced the introduction of its education reform initiative, the society engaged in an active, and at times intense, debate about this change. Reviewing media records was deemed as a necessary step to understand the reaction of the society to the new system and the role that the media might have played in advancing, or undermining, the reform. In order to arrive at a comprehensive and satisfactory understanding of the media's role and position it was important to review the publications of three major newspapers that are published daily in the country. It was also important to examine the recordings and contents of a popular local radio program that is broadcast everyday and deals with the concerns and complaints of the citizens. The radio program is widely regarded as an internal lens that keeps track of the reactions of the society to the affairs of the country.

The initial thinking was to conduct a random selection of 4-6 months every year starting with the 2004/2005 academic year, which witnessed the opening of the first cohort of reform schools. This required access to a lot of newspapers and radio programs records that the researcher knew, from her working experience, could not be easily obtained from their sources. The department of communication at the case study organization has the functional responsibility for monitoring, analyzing, cataloging and archiving all relevant media records. Working closely with staff at the department of communication, all newspapers and radio programs records were pulled out of the department's archive to examine the full universe and decide on the scope and size of the sample. A media record at the department of communication represented a copy of a newspaper article/editorial or a summary of the key topics that were discussed during a given episode of the radio program.

The key challenge that prohibited conducting a random sample was the unavailability of a comprehensive archive that included full daily records for all three newspapers and the radio program. For example, complete daily media records were not available for all months. Moreover, most of the records for the year 2006 were missing. This situation resulted in the selection of a review sample that was

based on the availability of daily media records for both local newspapers and the radio program. Only the months that had complete daily records were included and examined.

As shown in Table 4.2, media records for the first reform academic year (September 2004-June 2005) were fully included. The year 2006 was not included because there was not a single month in that year that had its full daily records archived. All the months in 2007 that had complete daily media records were included (January, August, September, October, November, and December). The first five months of 2008 were included; the end of May 2008 marked the last media records that were included in the analysis.

Table 4.2 also shows the number of newspapers and radio program records that were examined for each month of each year. In cases where the same piece of news appeared in more than one newspaper, each incidence was counted as one separate record. The table also shows the general position of the examined media records towards the reform. A judgment was made by the researcher as to whether any media record was negative towards the reform, positive, or neutral. The judgment was based on the content and tone of the record and the level of its support to the ideas of the reform or practices of the case study organization. A media record that had both negative and positive contents was counted as neutral. A media record that included announcements about the case study organization' activities or events was also counted as neutral. Appendix 1 includes a full bibliography of the examined media records. The objective of categorizing media records into positive, negative, or neutral was to get a general idea about the media's overall position and its role in endorsing or opposing the reform effort.

The researcher made copies of each media record for this case study research database. Each copy was marked by the name of the newspaper or the radio program and the date of publication or broadcast. All copied media records were organized by date and kept handy for further review. Indeed as the research study progressed, there were numerous times when the researcher pulled out these records and re-

examined their content to check against findings from the data that was collected through personal interviews and focus group.

Each copy of a media record was read at least twice before passing a judgment on its content as negative, positive, or neutral. The researcher also wrote down notes about positive and negative records in a separate notebook to facilitate future analysis.

Month	Year	Number of examined articles & programs	Negative	Positive	Neutral
September	2004	17	5	3	9
October	2004	9	2	3	4
November	2004	7	1	1	5
December	2004	12	0	6	6
Total 04 records		45	8	13	24
January	2005	12	3	1	8
February	2005	7	2	0	5
March	2005	15	4	2	9
April	2005	25	10	5	10
May	2005	39	24	3	12
June	2005	38	22	4	12
Total 05 records		136	65	15	56
June	2007	89	38	9	42
August	2007	19	6	1	12
September	2007	61	27	6	28
October	2007	58	19	5	34
November	2007	84	24	7	53
December	2007	51	15	13	23
Total 07 records		362	129	41	192
January	2008	79	27	12	40
February	2008	88	25	14	49
March	2008	119	22	35	62

April	2008	85	18	17	50
May	2008	95	22	17	56
Total 08 records		465	114	94	257
Total records		**1009**	**316**	**164**	**529**

Table 4.4: Summary of Examined Media Records

Prior to commenting on the media records review, it is important to draw attention to a caveat related to the comprehensiveness of these records. The selection of the media records was based on their availability at the communication department archiving system. It is quite likely that important observations were missed because of this selection method. The fact that media records from 2006 were not included should be taken into account when reviewing the results of the media review. Also, there is a possibility of bias as a result of the months that were examined. For example, there were no records examined for the month of July. The month of August is also poorly represented compared with the rest of the months. The researcher underscores that the following discussion is general and limited only to findings from the examined records.

This research study examined a total of 1009 media records that included articles and editorials from three daily national newspapers and one popular radio program. The review revealed that approximately 31% of the examined records were negative towards the reform, 16% were positive, and 52% were neutral or of a balanced nature. This means that negative coverage in the examined records was almost twice as much as positive coverage. To be precise, negative records were almost 93% (152 records) more than positive records.

An interesting observation was that few writers switched back and forth between fully supporting the reform and strongly criticizing its very basis. Perhaps this was due to poor understanding of the nature and details of this new initiative. Another

observation was that even the most hostile critics of the reform regularly acknowledged some positive aspects of the initiative and some even expressed limited degree of support to the change. The review also showed that the case study was actively issuing news release about its activities. Most of the 'neutral' records represented announcements by the case study organization or coverage of their activities.

In analyzing the content of the examined media records, several issues surfaced as key concerns about the philosophy, direction, and implementation of the reform. To start with, there was enough evidence to indicate serious concerns about **the compatibility of the ideas of the reform with the culture and national heritage of the country**. In particular, there were concerns about the influence of imported Western ideologies on the country's identity and local beliefs. An article that was published in October 7th 2004 warned that new curriculum at the reform schools represented a deliberate, and malicious, effort to '*erase our glorious national and regional history from the minds of our children*'. Another article that was published in September 26th 2004 criticized a curriculum that '*educates our students about the fourth of July, the US national day, instead of our own national day*'. Concerns about religious beliefs and the national language were also evident throughout the reviewed media records. An editorial that was published in January 17th, 2008 reported the experience of a parent with respect to deteriorating standards and poor attention to the subjects of religion and national language at her child's reform school.

The review also showed that another major concern was related to **the financial incentive scheme** that was initially a component of the reform's system. Under this scheme, direct financial rewards were to be given to the individual or sub-group that was shown by the performance measurement system to be successfully managing a school. There were several negative references in the 2004 and 2005 media records to the topic of financial incentives. Most of these references strongly criticized the introduction of profit-making into the education sector and warned against turning school administrators into businesspeople who would focus on increasing their

earnings at the expense of providing quality education. A number of editorials drew some dark scenarios about the future of education in the country as a result of 'commercializing' education.

However, the researcher could not find enough evidence to validate the supposition about the causal effect of the media's position on the reversal of the scheme by policy makers. The financial scheme was put to an end during the second cohort of the reform (the 2005/2006 academic year) without providing any explanation of the reasons behind the decision.

The third major concern revealed by the media records review was related to **the preparedness of the system and the availability of qualified human resources**. This concern highlighted a number of logistical errors related to students' registration procedures, teachers' hiring and dismissing practices, number of students in a classroom, and education staff' training programs. This third concern was particularly prevalent in the radio program records where most cases and complaints were related to implementation and logistical concerns. There were also a large number of complaints received by the radio program from parents regarding the behavior or qualification of teachers and administrators at the reform schools.

4.2   The Primary Data Collection Process

The secondary data described in section 4.1 provided the kind of historical account, or anchor point, which was required to better understand and explain the context of case studies research. Yet, and as expected, there was no readily available secondary data that specifically addressed the questions posed in this research. Most of the required data had to be gathered exclusively for this research, table 4.3 presents a summary of the different data resources used for this research.

The primary data in this research study was gathered using a mixture of personal interviews and a focus group as data collection instruments. Specific justifications

for the use of personal interviews and focus group to collect data were presented in Chapter 3. The use of multiple respondents and multiple sources of data in a case study as well as triangulation by multiple methods and types of data allows for greater substantiation of constructs and hypotheses (Eisenhardt, 1989a). The raw data that was collected during the interviews and focus group is retained by the researcher as part of the confirmability audit technique suggested by Lincoln and Guba (1985).

Resource of Data	Type of Data	Number of Sources/Respondents
Media Review	Archival	1009 Articles, editorials, and radio program records between 2004 and 2008 (excluding 2006)
Published/Official Documents and Memos	Archival	17 Publications and reports
Personal Interviews	Primary	11 Interviewees
Focus Group	Primary	12 Participants

Table 4.5: Summary of Data Resources Used for this Research

The following two sections, 4.2.1 and 4.2.1, offer a detailed account of the process that was followed to arrive at a final list of key stakeholders whose perspectives are potentially of importance to the questions raised in this research. It is worth noting that the process of identifying stakeholders was reiterative in nature and started with the initial review of secondary data and the early attempts to understand and analyze the case study organization. As analysis of the case study continued to take place, further insights resulting from this analysis helped identified more stakeholders.

4.2.1 Identifying key stakeholders

Even though the importance of stakeholders is well recognized in the literature (for example: Bryson, 1995; Rowley, 1997; Neely et al, 2001; Ackermann and Eden, 2001; Dixit, 2002; Boyne et al, 2004; Bryson, 2004), little attention has been given to this issue from a performance measurement perspective (Wisniewski and Stewart, 2004). This research aims to understand the ways performance measures' outcome information is utilized from the perspective of the key stakeholders. In particular, it explores the main factors that influence and affect utilization of measurement data and information from the perspective of these stakeholders.

It follows that a careful identification of key stakeholders, those who are conceivably able to provide feedback that can inform the questions posed in this research, is a critical step that is part of achieving a satisfactory data collection process.

While this research is not concerned with the notion of managing stakeholders per se, the literature on this subject can help guide an effort to <u>identify</u> the <u>relevant</u> stakeholders and understand the kind of <u>demands</u> they are likely to make on the organization. Ackermann and Eden's (2003) draw attention to some overarching concepts that can help the process of identifying the set of stakeholders that are most relevant to an organization.

The first concept highlights the importance of recognizing that each organization has a unique set of stakeholders. These stakeholders are likely to react differently to any strategic initiative taken by the organization. Indeed the reform can be considered a major strategic initiative that can potentially ignite different reactions by different stakeholders' groups. The issue about multiplicity, and hence heterogeneity, of stakeholders, as well as the relationships between these stakeholders, is also at the heart of any discussion about public organizations. This concept then applies to, and should be considered, in a discussion about the stakeholders of the case study organization.

The other concept suggested by Ackermann and Eden (2003) underscores the merit of paying attention to the relative importance of the stakeholders. Different stakeholders have different agendas and while it is important to identify these stakeholders as a first step it is equally important to recognize the differences in their agendas and potential roles which may range from supporting to sabotaging the successful realization of the set strategy and goals. This concept played an important role in directing the effort of charting identified stakeholders on the Power Interest Grid and was particularly useful in thinking through their different roles in the reform.

The process of identifying a list of key stakeholders for the case study initially required a review of the organization's mandate and an analysis of the groups and entities that are likely to affect and be affected by the strategies and goals of this organization. Initial reform proposals, yearly reports, publications, and articles that appeared in the media were used to highlight a list of stakeholders who are referenced in these records.

Fully aware that it was probably unrealistic to expect that all players with power could easily be detected, the researcher took the view that a key stakeholder can be any group or organization that has demonstrated any degree of power to directly or indirectly affect the delivery of the organization's strategy. The researcher depended primarily on the archival records review to identify all these stakeholders who, at least on record, seemed to have that feature. A key stakeholder could also be any group or organization that has enough interest in the strategy of the organization; regardless of whether or not they are likely to be affected by the strategy. As the following discussion reveals, some stakeholders, such as the media, demonstrated a lot of interest in the case study's performance measurement system even though they are not directly affected by its adoption and implementation.

It should also be noted that the process of thinking through a list of potential key stakeholders, albeit deeply anchored in documented records, mandates, and publications, was done judgmentally and intuitively based on a mixture of

documented evidence as well as the researcher's working knowledge and experience in the field. As an additional step, and in order to cross check the researcher's thinking on the subject, all interviewees were indirectly asked to think about a list of possible stakeholders. Two questions invited interviewees and focus group participants to think about organizations/groups they thought were likely to be interested in and/or affected by the performance measurement data produced by the case study organization. The responses to these questions showed that there was no need to make any changes to the list of identified stakeholders.

In arriving at a list of potential key stakeholders, the **national leaders** who introduced the education reform surfaced as one of the most apparent key stakeholders' groups. After all, the top-down reform was almost single handedly designed and initiated by these two leaders; it is only logical that they constitute a key stakeholder to the case study organization.

The list also included the **primary consultant group** that was charged from the very beginning with analyzing the education sector, developing several strategic reform options, and finally designing and implementing the adopted strategy. **Senior members of the governing board**, the policy making group that oversees the public organization under study, can also be identified as key stakeholders that directly affect the strategy and goals of the public organization. The **directors** who were appointed by the country's leaders and charged with running the newly established Assessment and Operational Entities, along with other members of the **management team** that report directly to these directors, were also identified as key stakeholders. This research refers to this group as the **senior management cadre**.

In addition to the above stakeholders, there are a number of external entities that can also be identified as key stakeholders. Teachers, schools' administrators, students and parents are clearly interested in the adopted change in the education sector and the outcomes of this change. The interests of this group, referred to as the **schools' system**, conceivably carry similar traits and focus more or less on the same issues and concerns. The **local media** can also be identified as a key stakeholder that, at

several points, is believed to have caused, or at least influenced, some of the changes in strategy that took place during the implementation process. The last identified key stakeholder is the community or **public** at large. A summary of the key stakeholders groups that were identified for this research study is included in table 4.4.

	Key Stakeholder Group
1	National Leaders
2	Primary Consultant Group
3	Governing Board
4	Senior Management Cadre
5	The Schools' System
6	The Local Media
7	The Public

Table 4.6: Final List of Identified Key Stakeholders Groups

4.2.2 Key stakeholders on the power interest grid

The identified key stakeholders for this case study represented an expected myriad of backgrounds, orientations, preferences, and positions. Based on archival evidence it was easy to infer that these stakeholders maintained different positions that varied widely from complete buy in and support of the reform initiative to adamant opposition and even attempts to stop the initiative. A closer look at these stakeholders revealed rather dissimilar levels of interest in the reform's strategies and operations. Similarly, these stakeholders enjoyed varied levels of power over the direction, and even destiny, of the reform. The following discussion serves to provide more details about the different stakeholders and where they can be charted on the Power Interest Grid.

The **national leaders** who commissioned the introduction of the reform have demonstrated great support to the reform's set strategies and goals and are plotted as

'**Leaders/Context Setters**' on the Power Interest Grid (Figure 4.1). The national leaders enjoyed the most power of all key stakeholders and shared the same vision with the **primary consultant** group that was charged with the initial design and implementation of the reform. However, the national leaders were not involved in the short term progress of the reform initiative and their interest level in following up on a number of operational issues could be moderate compared with, for example, the senior management cadre who worked at the case study organization. Moreover, the national leaders were occupied with the affairs of other sectors of the government; their interest could not be education focused only. Upon reflecting on where the national leaders should be positioned on the grid, the researcher decided not to place them at the high end of interest.

The **members of the governing board** were apparently supportive of the reform's set strategies and goals. Composed of senior and high level members of the public and private sectors, the governing body found it hard to convene at times due to the busy schedule of its members. This group of stakeholders enjoyed considerable power, although slightly less than that of the national leaders. They were also likely to have a lot of interest in the long term affairs of the reform, but not enough focused interest in its short term affairs to warrant placing them very high on the interest dimension. The governing board was also plotted as '**Leader/Context Setter**' on the Power Interest Grid.

The **senior management cadre** consisted of two directors who were appointed by the governing body and charged with presiding over the newly established Assessment and Operational Entities. The directors soon joined the supporting lines after an initial orientation and learning process. The senior management cadre also consisted of senior managers who held high level positions within the two entities and reported directly to the two directors. This group of key stakeholders was not necessarily homogeneous as each of the two senior directors, and consequently the senior management teams that reported to each of them, have different agendas and priorities that followed the mandates and functions of each entity. However, it was safe to assume that all members of this stakeholder group were heavily involved in

the affairs of the reform and the case study organization and that they enjoyed considerable power and interest. The senior management cadre group was plotted as '**Players**' on the Power Interest Grid.

The **primary consultant** organization that designed the reform had a vested interest in supporting it and promoting its goals. The primary consultant started out with a more active role in implementation but over the years handed over these responsibilities to the local staff. However, this group still enjoyed considerable power and interest that justified plotting them as '**Players'** on the Power Interest Grid.

Members of **the schools' system** were plotted as '**Subjects**' on the Power Interest Grid. This group consisted of **school administrators**, **teachers**, **parents,** and **students**. This group was believed to have a great interest in the outcome of the education reform initiative; even in its input and processes. However it was the researcher's judgment that the level of this group's power over the direction and strategy of the case study organization was not as significant as some of the other groups.

One reason that explained why the school system was plotted on the low dimension of power was related to the nature of the relationship that existed between the case study organization and some members of the schools' system, namely teachers and administrators. Through its authority over schools' employment policies and staff selection criteria, the case study organization was a party with considerable clout over teachers and administrators. Therefore, it was not difficult to appreciate how this group can be reluctant to exercise much power over the reform in general and the case study organization in particular. As for parents, many of them have little or no choice but to keep their children within the current education system that was being subjected to reform, even if they had concerns about the reform initiative. By the same token, students were not seen as a group that had the potential to influence the strategy of the public organization under study.

The **public** at large probably had a certain interest in the reform underway given the importance of the sector and the high stakes it involved: the future of the country's young generation. However, and due to lack of readily available primary or secondary data on the subject, the public's overall interest in, and power over, the public organization was difficult to estimate or even understand. This was not to suggest that the public's role should be undermined or neglected; rather this was to argue that any influence exerted by the public, in support of or against the reform, was likely to be indirect in nature and manifested through certain interest groups such as parents or the media. These groups were already identified as stakeholders groups and plotted on the Power/Interest Grid. The main interest of the public probably stemmed from the prevailing conservative nature of the society and a desire to preserve religious and traditional norms at any cost. Archival evidence indicated that the agenda of this ambitious reform was not necessarily perceived as in line with these norms. The public was plotted as low-power '**Crowd**' on the Power/Interest Grid.

As demonstrated in section 4.1.3, the local **media** has played an active role since the inception of the reform in drawing the public's attention to the practices of the case study organization in particular and the principles and direction of the reform in general. Indeed, the results of the media review provided a better understanding of the media's intensive engagement in the policies and strategies of the case study organization on almost a day-to-day basis. The observation in section 4.1.3 about the media's opposition of the performance incentive scheme serves as an example of particular relevance to this discussion. Tenacious and continuous attacks on the ethical underpinnings of one of the reform's key components, the performance incentive scheme, seemed to have ended with sudden abandonment of this component by the case study organization. Although the media review that was conducted in this research could not prove the influence of the media on the reversal of the policy, this example called attention to the possible power of the media. The media, as a stakeholder group, was expected to enjoy a level of power over the course of the reform that was close to that of the Players and Context Setters.

At the same time, the researcher found no evidence to support the notion that the media had a strong interest in promoting a certain educational philosophy or favoring a given direction of education reform. In fact, the media review indicated that most critics expressed some degree of support to the general ideas of the reform even if they have reservations about certain aspects of its programs.

The media group of stakeholders was plotted as high-power **'Crowd'** on the Power/Interest Grid.

Apparently, the identified stakeholders groups enjoyed different levels of power over the reform. The discussion also pointed out to the possibility that different stakeholder groups were likely to have diverse expectations and requirements of the reform. As a result, stakeholders groups were likely to have different level of interest in the divergent aspects of the reform's performance and, as a result, the performance measurement information that they would like to receive. Perhaps it was safe to assume that those who were directly involved in the original design and, later on the implementation of the reform, i.e. the "Leader/Context Setter" and the "Players", were expected to be more supportive of the reform and its strategies compared to the 'recipients' of the reform, i.e. the "Subjects" and the "Crowd". Figure 4.1 reflects the researcher's attempt to map the identified stakeholders according to their estimated degrees of power to affect the delivery of the reform's policies and programs as well as their interest in that reform. It also helps determine the relative significance of the different groups.

Figure 4.1: Mapping Identified Stakeholders on the Power Interest Grid

(Source: Ackermann and Eden, 2003)
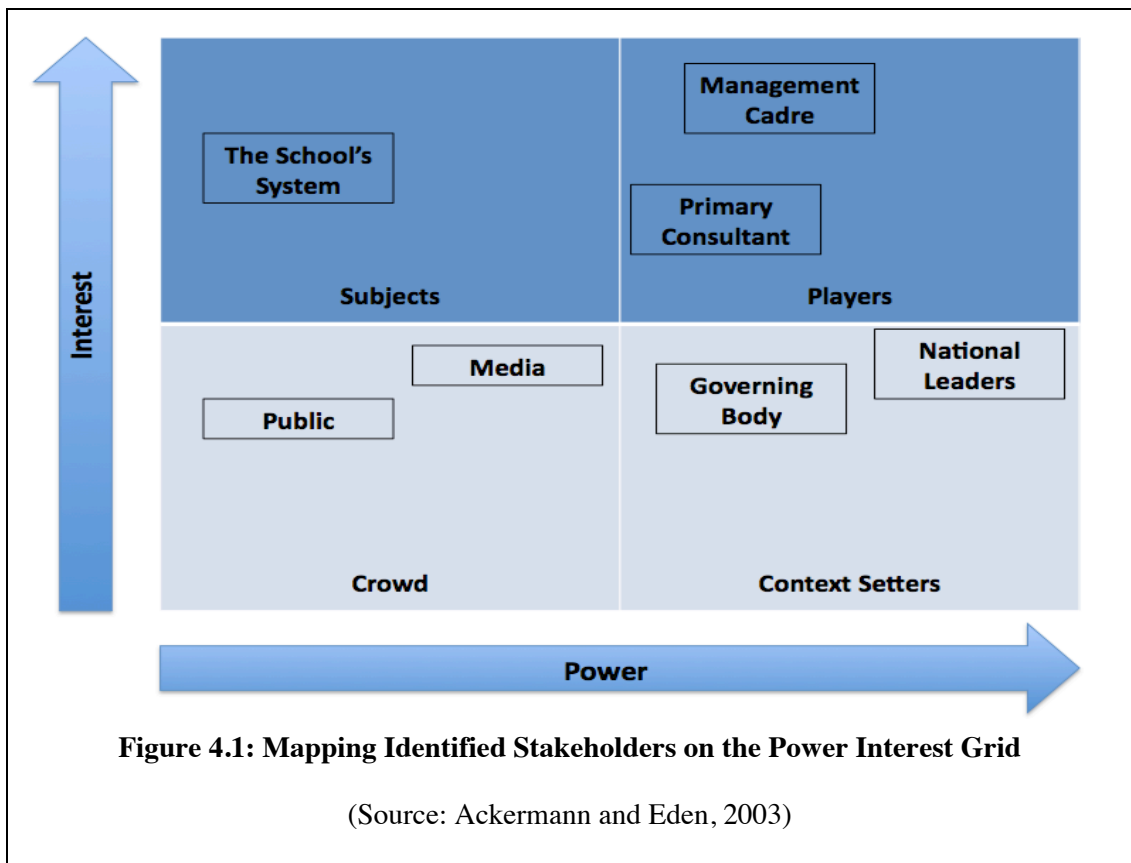
Upon identifying the key groups of stakeholders, and mapping these groups on the power-interest grid, the next step was to identify representatives from each group to start the data collection process. Justifications for using the selected data collection methods were described in Chapter 3. Table 4.5 provides a summary of the stakeholders groups that were represented in the data collection effort, the methods used to collect data, and the number of representatives from each of these groups.

The researcher conducted in-depth, personal interviews with one member from the governing board, two directors and four managers from the senior management cadre group, two staff members from the primary consultant group, and two active journalists from the media group. There were two failing attempts to interview one of the national leaders. There were also no interviews with representatives from the public stakeholders group. In addition, the researcher conducted a focus group with three schools' administrators, four teachers, three parents, and two students for the

schools' system group. The next section describes the primary data collection process in more details.

Stakeholder Group	Method	Number of Interviewees	Number of Participants
Governing Board	Personal Interview	1	
Senior Management	Personal Interview	6	
Primary Consultant	Personal Interview	2	
Media	Personal Interview	2	
Administrators	Focus Group		3
Teachers	Focus Group		4
Parents	Focus Group		3
Students	Focus Group		2
TOTAL		11	12

Table 4.7: Summary of the Number of Representatives from the Stakeholders' Groups in Each Data Collection Method

4.2.3 The personal interviews

Based on a comprehensive literature review, and the research questions that followed this review, the researcher developed an initial list of interview questions. This research is interested in exploring the ways outcome information, or data produced by the performance measurement system, is used by stakeholders, and, the factors that affect this utilization process. It follows that in thinking through the development of encompassing questions that address all components relevant to this research, it was necessary to envision the entire utilization process by tracking all the different phases that might follow the dissemination of outcome information.

Consequently, the interview questions were divided into three main categories. The first category of questions focused on **the dissemination process of the measurement data** and inquired about its quality and the formats used to deliver this data to stakeholders. The second category of questions focused on **the utilization process and the ways the measurement data was used**. Questions examined whether stakeholders used measurement data in any of the ways listed in the literature such as decision making, reporting, resource allocation, planning and monitoring. The third category of questions addressed the **factors that affect data utilization and examined the relevance and applicability of rational and political factors listed in the literature**.

The literature review provided the theoretical propositions that guided the development of the interview questions. Based on the literature review, a prompt list was developed which included several characteristics and factors that have been found to influence the success of performance measurement systems and the utilization of their measurement data. The prompt list was later on incorporated into the interview questions. Open-ended questions were followed by questions focused on a prompt list of possible factors.

The researcher asked two of her colleagues to review the interview questions and comment on their clarity and flow. Few minor editorial changes were made to the questions based on feedback received from the two colleagues. The revised list of interview questions was then shared with the researcher's advisors for their feedback and approval.

Following the advisors' approval of the questions, the next step was to conduct a pilot study. Three employees from the case study organization were selected for the pilot study. The wordings of two questions were slightly changed based on received feedback. One of the employees was mainly concerned with the suggestive tone of the question about the quality of the national education system. The question was reworded to ensure a more neutral tone: 'What do you propose to improve the system' was replaced with 'Do you have any ideas/ proposals to improve the

system'. Another question about the availability of procedures or systems to govern the utilization process was also slightly changed to address another concern about the clarity of the question.

The final list of interview questions included eighteen, mostly open-ended, interview questions (Appendix 2). For twelve of the eighteen questions, a number of sub-questions were added that focused on the prompt list of factors. A note was added after each question to remind the researcher to probe responses to the prompt list in order to ensure understanding. Probing also aimed at ensuring that the interviewer received enough insights into a given issue and collected specific examples when possible. As discussed earlier, the prompt list used in the interview questions represented the theoretical propositions for this research and were later on used as the research's priori codes during the data analysis phase. The prompt list ensured that theoretical propositions were well addressed during the data collection process.

A detailed interview protocol was developed to ensure a standardized and rigorous process is followed in all interviews (listed in Appendix 2). An interview protocol also helps guide and discipline the process of interviewing. The researcher mainly used her prior interviewing experience in thinking through and developing the interview protocol. A measure to establish construct validity was incorporated into the interview protocol by asking key informants to review transcripts of their interviews before the researcher conducted the data analysis.

Each interview was scheduled well in advance. The researcher would contact the office of each interviewee at least two weeks in advance and ask for an appointment after reading out a standardized form that has a brief description of the research, name of the researcher, academic affiliation, and the purpose of the meeting. A frequently asked question was related to the expected length of the interview and the answer given was 2-3 hours. Scheduling such lengthy appointments proved to be quite a challenge especially with some of the higher level stakeholders. The option of having the interview done over a course of two or more meetings was given in these cases. However, only one of the eleven personal interviews ended up being

conducted over the course of two meetings. This was the one with the member from the case study' governing Board. In the researcher's assessment, the fact that the interview took place over the course of two meetings had no serious implications on the quality of the collected data. The first meeting strictly covered the first group of questions about the measurement data. The second meeting covered the utilization process and the factors affecting that process.

A day before the interview, the researcher printed out a cover letter with the next day's date, the initials of the interviewee's first and last name, and the place of the meeting. This cover letter was stapled to the interview questions list as well as additional blank sheets of paper. This was done to prevent loss of papers and future confusion about the information source. On the day of the interview, the researcher arrived early with the oral consent form, and the stapled cover letter, interview questions and sheets of paper.

The researcher started each interview with reading the oral consent form. This helped introduce the research topic to the interviewees in a factual manner, announced the academic affiliation of the researcher, and reassured the interviewee that data will be used strictly for research purposes that are totally unrelated to the job of the researcher. The researcher's professional affiliation was feared to potentially influence at least some interviewees' willingness to divulge information. The researcher believes that the disclosure of the exact goals of the interview, and how shared information will be used, did help alleviate some of the initial apprehension about the purpose of the data collection effort.

Another important purpose served by the oral consent is the promise to keep the identity of the interviewee anonymous to the extent possible. The highly politicized environment of the case study, as well as prevailing cultural attitudes about sharing information and opinion, made it significant to address issues related to privacy and confidentiality. Prior to starting the interview process, each interviewee was asked to share any concerns or feedback they had at that point. This was a precautionary step to address any issues that the researcher did not account for. Indeed six out of

the eleven interviewees had additional questions to ask about the academic affiliation of the researcher, the nature of the graduate program, where the results of the research will be published, and how the research effort was being funded. Many interviewees asked for a copy of the final research document when it's ready.

During the interview process there was no use of recording machines. Based on working experience in the country, and bearing in mind the general cultural uneasiness towards recording conversations during which sharing sensitive information takes place, it was the researcher's judgment that avoiding the use of recording machines was a safer course of action. Hand written notes were used to record the responses in the space provided underneath each question on the interview form and the extra sheets of paper.

Most of the interview data was captured through making detailed linear notes as the interview was in progress. The researcher sometimes jotted down special notes that are not necessarily related to the questions being asked or the answers being given. For instance one of the interviewee changed the course of her answer after pausing and thinking quietly for few moments. The notes indicate that the first few words that came out of that interviewee's mouth were complimentary and almost official in tone. The interviewee then changed to a more critical tone and started sharing a somewhat controversial point of view. The researcher's judgment was that this change of position should be noted on the response sheet. It should also be added that all of these additional notes were later on added to the research's Coding Sheets.

Almost all interviews went quite smoothly with interviewees feeling comfortable about responding to questions and expressing their opinion as evidenced by the positive reactions of the interviewees. In fact almost all interviewees were quite enthusiastic about the opportunity to comment on the reform and the use of performance data. One exception was one of the senior management cadre interviewees who expressed his lack of appreciation for research in general and his disbelief in its capacity to change real life situations. This interviewee was quite resistant to getting engaged in a full-pledge discussion about the different topics

covered in the interview, especially at the beginning of the interview. The gradual improvement in his attitude over the course of the interview allowed the researcher to return to some of the earlier questions for more clarification. However, the researcher noticed that interviewees were apparently very interested in the subject and were readily available to share a lot of insightful and useful information, especially after initial concerns about confidentiality were addressed.

On average, interviews lasted for three hours each. The longest interview lasted for four and a half hours and the shortest one for one hour and thirty minutes. The researcher kept track of the length of the interviews by using a timer that was set immediately before asking the first question. Reported time was rounded to the closest quarter of the hour. A total of eleven in-depth, personal interviews were conducted over the course of seven months (November 2007- May 2008). All interviews were transcribed into Microsoft Word documents. Almost all of the transcriptions were made within 24 hours of the interview day to ensure fresh recall of events. However it took the researcher up to 48 hours before transcribing two of the interviews into a Word document. All interview transcripts were printed out to facilitate reading, recording notes, and highlighting certain responses using color-coded themes. The researcher also noted down any missing information, unclear answers, and other points that need further clarifications or data collection.

One of the measures suggested in the literature to establish construct validity includes asking key informants to review transcripts of their interviews before the researcher conducted the data analysis (Miles and Huberman, 1994). A draft interview transcript was sent to each interviewee filled with his/her own responses; all interviewees were asked to validate their responses. Immediately after sending the transcript, the researcher would call and ask for an appointment to discuss their comments and feedback as well as ask additional questions if necessary.

Three out of the eleven interviewees agreed to conduct a second interview to communicate their feedback and answer questions; six agreed to have a phone conversation; and two interviewees gave the researcher their consent regarding the

draft transcript. Each of the three second interviews lasted for about an hour; a notably short time length compared with the first round of interviews. The six phone conversations were also quite short but served the purpose behind conducting them. For one of the two interviewees who had no feedback to provide, the researcher requested a phone meeting to ask for some missing information. The interviewee accepted to conduct a phone meeting during which he supplied the researcher with the missing information.

After validating responses, and asking additional questions when needed, revised interview transcripts were produced. The process of extracting data from the final interview transcripts into the interviews' Coding Sheet started immediately after. Details of the coding and analysis processes are covered under Data Analysis in section 4.3.

4.2.4 The focus group

The in-depth personal interviews that took place over a period of seven months allowed the researcher to deeply explore the questions put forward and get detailed information about a complex and highly sensitive subject from several identified groups of stakeholders. This left the schools' system which was earlier on identified as a key stakeholder group for this research. This group of stakeholders consisted of schools' administrators, teachers, parents, and students. As discussed earlier, the schools' system group was quite distinct in its characteristics and level of power compared to other stakeholders covered in the personal interviewing process. It was also likely that this group collectively harbour more or less corresponding concerns about the education reform given their proximity to the education process and their intimate, day to day, interactions with parallel components of the school system. Moreover, a joined perspective of this particular group can be more relevant to this research given the fact that these were heavily populated groups. While conducting a personal interview with one out of six senior members of a governing board might be considered good enough to get the perspective of the board, conducting personal

interviews with one out of a thousand teachers might not necessarily provide the research with the required broader perspective of the group.

A key advantage of using focus groups includes provision of rich data. Focus groups also provide the opportunity to stimulate respondents' thinking and recall aiding, cumulative and elaborative (Denzin and Lincoln, 1994). These are unique characteristics that are not necessarily evident in personal interviews but certainly complement that data gathering technique. Consequently, it was the researcher's judgment that using focus group would allow for another level of data gathering or perspective on the research questions not necessarily available through individual interviews

Due to cultural considerations related to the gender of a person seeking access to schools, the researcher selected a girls' school. The segregated local school system does not welcome visitors from opposite genders inside the schools' walls. The selected school was part of the first batch of schools that underwent the reform under study and was in its fifth year of operating as new reform school when the focus group meeting took place. Such history was thought to have increased the participants' knowledge of the performance measurement system and their ability to provide feedback that was more likely to be based on real experience rather than expectation and/or prognosis.

The school was approached with a request for a focus group meeting two months before the meeting took place. The researcher contacted the principal over the phone and read the standardized form that has a brief description of the research, name of the researcher, academic affiliation, and the purpose of the requested meeting. The principal was requested for a group meeting that includes administrators, teachers, parents, and students. The researcher also proposed that the total number of participants in the meeting is kept under 15 people.

The logistics of getting the different people together at the same time was quite challenging. The calendars of potential respondents did not match and the fact that

this meeting took place toward the end of the academic year made scheduling even more difficult. The researcher kept coordinating with one of the school's administrators until a mutually agreed upon time was arrived at two months after the initial request. It is worth noting that the procedures listed in the interview protocol were strictly followed throughout the process.

A day before the focus group meeting, the researcher printed out a cover letter with the next day's date, noted down that this was the focus group meeting, and wrote down the name of the school as the place of the meeting. The cover letter was stapled to the interview questions list as well as additional blank sheets of paper. The names of the focus group participants were not shared with the researcher up to that point.

On the day of the focus group meeting, the researcher arrived early at the school with the oral consent form, and the stapled cover letter, interview questions and sheets of paper. The researcher was escorted to a large meeting room where the meeting was going to take place. The focus group' participants started arriving and were introduced to the researcher who took notes of their names on a separate piece of paper to facilitate communication and make it easy to call them by name.

The total number of participants was 12 people; one principal, two administrators, four teachers, three parents and two students. The focus group meeting started with the researcher reading the standardized oral consent form. Participants were introduced to the research topic and the academic affiliation of the researcher was announced. The intended use of the collected data was also explained and participants were promised that their identity would remain anonymous to the extent possible. Two of the parents responded that issues of privacy and confidentiality were not of any concern to them and that they were willing to share all their thoughts without any reservation. They asked the researcher to start asking questions. However, and in adherence to the interview protocol, the researcher asked the participants if they had any questions or concerns before starting the meeting.

The focus group meeting lasted for four hours and fifteen minutes excluding a ten minutes break after the first three hours. The same eighteen questions that were used during the personal interviews were used during the focus group. There was no use of recording machines for the same reasons given under the personal interviews discussion. Hand written notes were used to record the responses mainly on the blank sheets of paper. The researcher marked blank papers with the English letters A, T, P, and S. Answers from the administrators group was placed on the blank papers marked with 'A', from teachers on those marked with 'T', from parents on those marked with 'P', and from students on those papers marked with 'S'.

The interactions were very intense in the sense that all participants were quite engaged and wanted to share their points of view (often at the same time). Managing such a diversified, and quite large, focus group required a lot of work. Interactions and inquiry needed to be well directed and systematic questioning of several individuals was taking place simultaneously during that single setting.

In her professional capacity, the researcher had conducted tens of focus groups over the past five years in the same country of the case study organization. This working experience, and its acquired skills, proved to be very helpful in managing the meeting and preventing a small coalition of participants from taking over the discussion and influencing the answers of the remainder of the group. This coalition was led by one of the teachers who actually tried to stop others from talking and at one point ridiculed an answer that was not to her liking. The researcher made every effort to encourage 'all voices' even if they did not fit into the "group think". At several points, the researcher started assigning turns in an attempt to provide all interviewees the opportunity to express their ideas. Moreover, the researcher would pose after each question has been addressed and turn to silent participants inviting their input. It can be stated that all participants were provided with a fair and equal chance to share their perspective and feedback for all questions.

Throughout the course of the focus group meeting, the researcher was particularly concerned with 'getting it right' the first time. Knowing in advance that this was

likely to be the only focus group that would take place, and that as a collective interview it would not make sense to send the interview transcripts back to all participants to validate the answers, the researcher exerted every effort to make sure all required data was collected and there were no missing or unclear answers. Special attention was given to ensure every probe was explored, every sub-question was asked, and at least one representative from every sub-group provided some kind of feedback. The researcher also jotted notes to indicate cases where participants stated they have no comments.

Despite the quite long duration of the focus group, the researcher did not have the opportunity to write down the full and complete answers in all cases. In fact, most of the notes that were jotted down by the researcher during the focus group were short and not complete. Over the next couple of days that followed the focus group, the researcher completed these short sentences and ensured that the full sentences replaced brief notes.

Within the 24 hours that followed the focus group, all notes were transcribed into a Microsoft Word document. More details to short and incomplete notes were added over the following 24 hours. The focus group was conducted during the first week of June 2008. Immediately after the focus group transcripts were finalized, the researcher started to extract data from these transcripts and stored them into the focus group' Coding Sheet. A detailed description of the coding and analysis stages is provided in the following Section.

4.3   The Data Analysis Process

Even prior to commencing data collection, there were several initial attempts to design Coding Sheets that can be used during the following data analysis phase. Effort to develop Coding Sheets continued throughout the data collection phase. A number of versions were produced on Excel until the fifth version was considered final and ready to be used for data analysis. Initially, the research questions were

vertically listed on the left side of the Excel Sheet following the same order on the interview questions form. Corresponding priori codes were then listed to the right of each question. As discussed before, all the priori codes for this research were included in the interview and focus group questions as part of the prompt list.

After several editions, and in an attempt to infer more comprehensive and overarching themes from the codes, all priori codes were placed to the left of their corresponding interview questions and further clustered into an initial list of key topics that were placed to the left of the codes. Enough space was left under each priori code to allow for additional codes that might emerge out of the data analysis stage. Figure 4.2 shows a partial image of the final version of the Coding Sheets. The codes in red denote an emerging theme. The flow of the final Coding Sheet cascaded from very general themes, or key topics, to the priori codes, to the research questions. Two separate Coding Sheets were produced: one for data from interviews and another for data from the focus group.

Each respondent was assigned a numerical code at the top of each of the relevant Coding Sheet. In order to address concerns of confidentiality, each respondent was assigned a code that starts with the letter R (for respondent). Information regarding the identity and affiliation of the respondents was stored in a separate Microsoft Document. R1 to R11 denoted respondents that participated in the personal interviews. In assigning serial numbers for the interviewees, the researcher purposely placed respondents from the same key stakeholder group next to each other on the Interview Coding Sheet. The goal was to facilitate the detection of systematic patterns or relationships within and across stakeholders later on during the data analysis phase. Likewise, each participant in the focus group was assigned a numerical code (R12 to R23). Responses from the schools' administrators and principals group were added together under a column marked 'R12-R14' on the Focus Group Coding Sheet. Responses from the teachers group were added together under a column marked 'R15-R18'. Responses from the parents group were added together under a column marked 'R19-R21', and finally responses from the students group were added together and placed under a column marked 'R22-23'.

Interview Coding Sheet
(Version V)

Key Topics ↓	Codes ↓		R1 (1hr 45mnts over 2 days)	R2 (3hrs 30mnts)	R3 (4hrs)
		Q1 Receive data from the PM system. What data	Yes, results of educati	Yes, school articles a	Yes, school articles, r
1. Timeliness		**Q4**			
	1.1. Set Schedules for Data Release	There is an estbalished timetable	No. We never know w	There is a timetable b	There is a timetable fc
	1.2. Concerns Over Continuous Delays		Data is chronically late	It's always too late to	There are factors outs
2. Format of Data		**Q5**			
		General data delivered in open files		National tests and school surveys are delive
	2.1. Open Files	Specific data delivered in open files		We deliver specific po
	2.2. Data Includes Analysis	Delivered data has at least some analysis		We urgently need to s	Performance data sho
	2.3. Data Includes Graphics/Visuals	Delivered data has at least some graphics/illustrations	National tests results and exuctive reports pi	When we share natior
	2.4. Part of a Report	Data is delivered as part of a report/standardized report	School Articles always	The School Articles ar	All of the School Articl
				T
		Other forms		
3. Changes		**Q6**			
	3.1. Stable Data	Data cover same topics each time	For the most part, the	Data categories have	Certainly not. Topics
	3.2. Changes Experienced	Data topics change from time to time			We have added many
		Change in covered topics is due to…			We always follow up c
		Change can be initiated by respondent	The Board has asked	My entity asked for many changes in the cov
4. Satisfaction with PM Data		**Q7**			
	4.1. Clarity	Data is clear	Data is clear but not a	No. Too much informa	Yes, I believe the data
	4.2. Organizion	Data is organized	Data included in the E	Except for the report c	It is organized althouc
	4.3. Relevance to user	Data is relevant to user	I don't find it very use	Not at all. Again, our r	The main reason data
	4.4. Focus	Data is brief and focused	What we get (at the B	No; masses of data is thrown at us
	4.5. Level of details	Data has enough detail	Yes, enough detail to	Too much details which is not even necessa
	4.6. Level of analysis	Level of analysis in data	Not adequate at all. A	No analysis at all. Thi	We should only collec
	4.7. Base line comparison	Compares with past data	Not always. I believe t	Not enough and not u	Our report cards have
	4.8. Visual illustration	Quality of illustartion	Can be much better	very weak	We try to incorporate

Figure 4.2: The Final Coding Sheet

In an attempt to organize and discipline the analysis process, the researcher consulted and used the framework suggested by Miles and Huberman (1994). The framework draws attention to three distinct stages of analysis: ***data reduction***, ***data display***, and ***conclusion drawing and verification***. In addition to benefiting from the systematic guidance of this comprehensive framework, the researcher constantly engaged in the 'constant comparison' technique suggested by Gall *et al*. (1966) during the data collection and data display stages. The iterative technique involves repeated examination of the data segments to determine if they were placed in the most suitable categories. It also involves revising, polishing, and modifying the categories if found necessary. The use of the constant comparison technique complemented the utilization of the Miles and Huberman framework and added further order and discipline to the data analysis effort.

4.3.1 Data reduction stage

During the initial stage of **data reduction**, work focused on organizing the mass of data by meaningfully condensing and transforming it in a way that served the salient research questions but also remained open to the possibility of inducing new meanings from the gathered data. The focus was on distilling what the different respondents suggested about the performance measurement system, how they viewed it, how they used its data measurement, and why. At the same time, the researcher had to balance the data reduction goal with deliberate effort not to compress the data to the point that they look like they were collected through close-ended questions. It would be a waste of effort and potentially valuable data to thin down the richness of the collected data.

As described under 4.3 above, the Coding Sheets were carefully developed to match and serve the research questions, make sure all priori codes were covered, and at the same time allow room for the addition of new meanings if applicable. Each interview transcript was read by the researcher at least twice before extracting its data and transforming it into these Coding Sheets. The several rounds of reading also aimed at familiarizing the researcher with all of the data before moving on with the analysis process.

The availability of a 'ready to fill' list of priori codes made the first analysis attempts a relatively uncomplicated, albeit not easy, undertaking since the researcher did not have to start out with searching the data and arranging it into broad topic areas. Table 4.6 shows an example of two key topics and their priori codes at an early stage of the data analysis. The table also shows that additional data that did not fit into any of the priori codes was listed under the 'others' column at that point. It soon became clear that quite a number of responses from the both the personal interviews and focus group did not fit into the priori codes. 'Dumping' all this kind of data in the extra space available for 'others' was not a viable or sustainable option.

To address the issue regarding data that did not readily fit into any of the existing priori codes, additional readings of the transcripts took place to ensure this data could not be placed in any of the other priori codes on the Coding Sheet. In some cases, responses given to a certain question could not be placed in the priori codes for the same question. Before creating a new emergent code, the researcher had to examine all other priori, and emergent, codes that existed on the Coding Sheet at that point to see if the response fitted well in any of them. When a matching existing code was found, special notes were added by the researcher to explain the movement of the collected data across the Coding Sheets.

In addition, the development of any emergent code was followed by a process during which each new code was checked time and again against the transcripts to make sure it captures the exact meaning listed in the collected data. The process of defining and arriving at a final list of emerging codes continued throughout the data reduction and data display stages.

Priori Codes / Topics	**Priory Code 1**	**Priori Code 2**	**Priori Code 3**	**Priori Code 4**	**Priori Code 5**	**Priori Code 6**	**Others**
Rational factors	Financial Resources	Human Resources	Technical Knowledge	Access to data	Clarity of goals and strategy	Organiz-ational set up	Yes
Soft factors	Cultural attitude	Internal interest groups	External interest groups	Role of the media	/	/	Yes

Table 4.8: Examples of Two Topics and Their Priori Codes during the Early Stages of Analysis

Throughout the process during which data segments were sorted into each code and new codes were being developed, the researcher was continuously engaging in the

referenced constant comparison technique suggested by Gall *et al.* (1966). The researcher repeatedly challenged her own decisions to place data in a given code and examined the data segment time and again to determine with more certainty that they were placed in the most suitable codes. Indeed engaging in constant comparison helped clarify the meaning and appropriateness of each code and created clear distinctions between the different codes. The technique was repeated several times until the researcher was comfortable that the creation of new codes was no longer necessary to account for the phenomena under study, that no new data emerged that was relevant to the established codes, and that the relationships among the different codes were well established.

A number of revisions took place as a result of applying the constant comparison technique. For example, at the beginning of the data reduction stage the researcher found a need to develop an emerging code titled 'internal coordination'. There was good evidence from the collected data that poor internal coordination within the case study organization was negatively influencing the utilization of measurement data. Respondents were not clear about the purpose of collecting performance data and, at times, were not even aware of the data collection effort in the first place. The development of a new 'internal coordination' code seemed to be well justified to cover emerging data segments that were relevant to the research salient question. Relevant data segments were placed in this new code for weeks. After several readings of the data that was placed in this emergent code, the researcher started to notice that this data can also fit well under the 'internal communication' priori code. The engagement in constant comparison between the two codes helped clarify the meaning and appropriateness of the priori codes and confirmed that there were no clear distinctions between the two codes to justify the separation of the data segment. After some deliberation on the matter, the decision to create the emergent code 'internal coordination' was reversed and all relevant data segments were placed under 'internal communication'.

Perhaps it is important to note that, in retrospect, the data reduction stage was the most difficult for the researcher. The idea of cutting down rich and seemingly all

important data so that it addresses the salient research question was in fact almost painful. Every collected word came from a real person and described real events that meant something to someone. However, difficult decisions had to be made regarding feedback that did not relate, or distantly related, to the main issue of the research, regardless of how captivating or otherwise interesting it looked.

The meaningful condensing and transforming of mass data that took place during the data reduction stage resulted in the development of 9 emergent codes and the final placement of all data segments in a total of 49 priori and emergent codes. A full list of priori and emergent codes is included in Table 4.7. While the initial development of priori codes was based on the review of the literature and borrowed from previously developed constructs or theories, the possibility of coming across a piece, or pieces, of collected information that might suggest an additional perspective or variable was well considered right from the beginning. This is to be expected in qualitative research which should remain open to inducing new meanings for the available data even though initial categorizations are typically shaped by pre-established study questions. Another important change that resulted from condensing and transforming the data was the revision of the key topics. While their number remained unaffected, there were few changes to the titles of these key topics to better reflect the developed mixture of priori and emergent codes.

The end result of this stage of the research was the arrival at a mix of codes that were derived both from previously developed constructs as well as from the data itself. In addition, the revised and final list of emerging and priori codes on both Coding Sheets were full with data segments that were directly extracted from the interviews and focus group transcripts and could be later on used as cited evidence. These excerpts were carefully selected from the interview and focus group transcripts to fit into each code. The data reduction stage was an ongoing process that span over a period of nine months (December 07– September 08). It started soon after the beginning of data collection, and as interview and focus group transcripts became final and ready for data extraction, and continued for three months after the conclusion of data collection.

Key Topic 1	Priori Codes	Emergent Codes
Timeliness	Scheduled release	Constant Delays

Key Topic 2	Priori Codes	Emergent Codes
Format	Open files	
	Data with analysis	
	Data with Graphs	
	Part of a report	

Key Topic 3	Priori Codes	Emergent Codes
Stability	Stable data	
	Changes experienced	

Key Topic 4	Priori Codes	Emergent Codes
Quality of PM Data	Clarity	
	Organization	
	Relevance to user	
	Focus	
	Level of details	
	Level of analysis	
	Baseline data	
	Visual illustrations	

Key Topic 5	Priori Codes	Emergent Codes
Quality of PM System	Data collection	Measures mix
	Data analysis	System origins
	Aligned with needs	Review process
	Aligned with goals	
	Technology use	

Key Topic 6	Priori Codes	Emergent Codes
Utilization	Monitor performance	Crisis management

	Evaluation	Misuse/inactive use
	Planning	
	Reporting	
	Operational decisions	
	Strategic decisions	
	Policy changes	
	Board meetings	
	Public briefings	

Key Topic 7	Priori Codes	Emergent Codes
Communication	Intra-organizational	External communication

Key Topic 8	Priori Codes	Emergent Codes
Rational Factors	Financial resources	Legal requirement
	Human capacity	
	Technical knowledge	
	Access to data	
	Clarity of strategy	
	Organizational set up	

Key Topic 9	Priori Codes	Emergent Codes
Soft Factors	Culture/Attitude	Political commitment
	Internal groups	
	External groups	
	Media	

Table 4.9: List of all Priori and Emergent Codes for the Research

4.3.2   Data display stage

The next step required thinking about methods and techniques to exhibit the data in a way that facilitated arriving at inferences, identifying relationships, and drawing conclusions. The goal of the data display stage according to the Miles and Huberman model (1994) is to provide '*an organized, compressed assembly of information that permits conclusion drawing…*' From the perspective of performance measurement utilization, data display was regarded as a potentially key tool in identifying and explaining why the measurement system was, or was not, working in a way that facilitates utilization, and, what might be done to address any existing challenges or improve the system. It follows that the researcher needed to carefully detect patterns of interrelationships throughout the data display stage to understand why the performance measurement system was working the way it was.

During this stage, the researcher depended heavily on the use of coloured highlighters and symbols as a visual method to extrapolate and discern patterns and interrelationships. For example, similar responses were colour-coded in the same shade of colour. Anomalous data segments, such few random references to favouritism and patronage, were colour coded in yellow with a question mark added to it. A combination of yellow and a question mark aimed at drawing the researcher's attention that further data search is required to check if similar patterns were existent elsewhere in the respondent's, or other respondents', data excerpts, under the same or other categories.

Moreover, an attempt was made to assign weights to the different codes and themes by counting the number of respondents who provided similar patterns of answers. Given the relatively small number of respondents, the researcher was fully aware that 'counting' similar answers did not necessarily provide evidence that one code or theme was more important that another. Rather, weights were merely used as general indicators, or signs, to highlight what themes were conceivably more widespread amongst the respondents.

Throughout the course of the data display stage, the researcher conducted constant comparisons among the pieces of collected data in order to identify patterns and relationships within each interview, within the focus group, across the different interviews, and across interviews and the focus groups. In addition, Gall's (1966) constant comparison technique of comparing and contrasting across instances was repeatedly used to first establish important patterns and then to question and refine these patterns as part of the ongoing analysis effort.

To facilitate work during this stage, several copies of each of the filled Interview and Focus Group Coding Sheets were printed out on A3 paper in order to allow the researcher to liberally scribble notes, draw lines, and use different colours and symbols to underscore different themes and meanings. On a separate A3 paper, the researcher posted Berkowitz's (1996) suggested six questions to serve as a constant reminder of what to look for in detected patterns.

While the boundaries between the data display and conclusion drawing stages proved to be a bit blurry in this research, some of Berkowitz's questions seemed more relevant to the data display stage and were more frequently used by the researcher during this stage. In particular, questions related to emerging common themes about each topic; the relationship between these patterns and the research questions; and any noteworthy deviations from these detected patterns were constantly asked by the researcher during the data display stage. These questions needed to be carefully pondered, and appropriately answered to the extent possible, at the data display stage versus the conclusion drawing stage. Put in other words, the verification of detected patterns, it can be argued, should start to take place as early as possible to facilitate the progress of the analysis to the final stage of conclusion drawing. Delaying posing these particular questions until the following stage, when pattern matching would mainly be used to reflect and enhance validity, seemed to be impractical if not detrimental to the expected progress of the analysis process.

A related point is that the researcher was naturally interested in examining data per group of stakeholders and comparing data across the different groups of

stakeholders. It was important for this research, from an analytic point of view, to uncover the relevant differences in perceptions among different groups of stakeholders. Such patterns could potentially provide valuable insights into who the performance measurement system was best serving and, hopefully, why. This means that there was a need to arrive at the 'right', or perhaps the best possible, answers to Berkowitz's questions as early in the analysis process as possible in order to allow for better, and more grounded, conclusion drawing later on.

In addition, Berkowitz asks a very relevant question regarding the effect of participants' environments or past experiences to their behaviour and attitudes. The researcher found that a more relevant question to ask in this research would be related to the participants' affiliations and positions on the Power/Interest Grid of stakeholders. Do participants' affiliations with a given group of stakeholder influence and shape their behaviour and attitudes? Is this true for all groups of stakeholders?

In order to address the question about stakeholders' collective thinking on the different issues, separate A3 working Coding Sheets were printed out and dedicated for this purpose. As explained earlier, the design of the Coding Sheet purposefully clustered respondents from the same key stakeholder group, as identified on the Power/Interest Grid, next to each other on the Excel sheet. For each code, the researcher would search horizontally on the Coding Sheet, across data from all respondents, for similar meanings. Detected patterns would be colour coded using different colours. The researcher would notate the numerical codes of the respondents who seem to carry similar views regarding the priori or emergent code in question. For example, respondents who gave answers that indicated they believed that the foreign origins of the performance measurement system, an emergent code, worked against its ultimate utilization were colour coded in red; respondents who did not reference this issue were colour coded in green. On the working sheet, narrative notes indicated the numerical codes of the respondents who agreed on this point and those who did not. Subsequent notes described patterns of cross-stakeholders similarities and differences.

The working Coding Sheets proved very helpful in highlighting patterns of cross-stakeholders group differences and similarities. At times, these patterns were so clear and could be visually detected on the sheet even prior to using other methods to bring them out. In general, there were strong intra-stakeholder group similarities for most codes. The data display found that participants' affiliations with their identified group of stakeholders did indeed influence their attitudes and perceptions on most examined issues. The finding was generally true across all groups of stakeholders except for the case study management cadre group. This group showed the most intra-stakeholder group dissimilarities. In particular there were differences among this group of stakeholders regarding how its members perceive the quality of the performance measurement system and the measurement data it produces.

In addition, data display showed clear overlaps across participants from all stakeholders' groups regarding a number of priori and emergent codes. One of the most important inter-stakeholder groups' similarities that surfaced during this stage was regarding the key issue of measurement data utilization. There were clear overlaps across all participants, from all stakeholders' groups, over their poor utilization of measurement data. This particular inter-stakeholder groups' similarity contributed to a major conclusion drawing about the performance system and later on shaped the course of the discussion about the findings as the next chapter shows. Another example of strong overlap was related to the dissatisfaction of different stakeholders group with the focus of the measurement data. Other examples included agreement over Schools' Articles as the most effective method to disseminate measurement data.

It is worth noting here that all analysis processes were conducted manually and without utilizing any of the widely available qualitative data analysis softwares. The researcher initially wanted to utilize Atlas software to help detect themes in the collected data. However issues related to translation made it difficult to move in that direction. Almost all of the interviews were conducted in Arabic. The researcher was instantly translating the responses and most of the notes were hand-written in

English during the interview. However, some notes were written down in Arabic because the response enclosed certain unique expressions or culturally-exclusive language. The researcher thought it would be better to keep these notes in their original language of communication to preserve meaning.

4.3.3   Conclusion drawing and verification stage

During the course of the final stage of **conclusion drawing and verification**, focus shifted back to the research questions and the implications of the analyzed data for these questions were examined and assessed. Verification of data necessitated revisiting the data time and again until the researcher was comfortable that the evolving conclusions can be substantiated. Pattern-matching was used again to further enhance validity. Drawing conclusions from data analysis in quantitative research should keep in mind concerns related to their credibility and defensibility. Meanings that emerge from the collected data should be "tested for their plausibility, their sturdiness, their 'confirmability' - that is their validity" (Miles and Huberman, 1994, p.11). The commonalities and differences between the data were compared, and the pattern of evidence was established, in order to support the claimed findings of the research.

The researcher also kept a record of the common and most referred to themes that emerged in the responses about specific topics, and how these patterns (or lack thereof) helped to illuminate the broader study questions. Deviations form these patterns were also noted and an attempt was made to explain these deviations. Moreover, the researcher noted few interesting stories that emerged from the data (Berkowitz, 1996). It was also during the conclusion drawing and verification stage that the researcher started thinking about reporting on the data findings in a way that was more in line with key emerging conclusions rather than the order of the research questions.

One of the challenges in analysing across the perspectives of multiple groups of stakeholders was not to fall into the trap of deciding which data were most accurate. Different perspectives were to be expected as a natural product of the differences in respondents' experiences, positions, and, as argued before, power and interest. The researcher was aware of the need to carefully knit all the perspectives together in a final report on findings that address the main focus of the posed questions. At the same time, the respondents for this research represented a mix of highly knowledgeable senior officials and others who were much less knowledgeable on the subject. For example, the researcher thought it was quite fitting to give responses from the policy making group greater weight in the analysis than responses from the students group for example.

As a reasonably coherent set of explanations started to get developed for the prevalent type of measurement utilization, more contrast and comparison, pattern and theme noting, clustering, and variables partitioning took place iteratively and simultaneously to draw and test the final conclusions. The next Chapter discusses these findings in great details.

CHAPTER 5

FINDINGS AND DISCUSSIONS

5.1   Introduction

In the previous chapter, descriptions of the case study organization, primary data collection processes, and data analysis processes were discussed. This chapter aims to report on five key findings of the study. These findings are reported in an order that does not strictly follow the order of the research questions listed in Chapter 3. The researcher found it more helpful to cluster the findings around key issues that were repeatedly highlighted by the interviewees, as revealed by the data analysis, rather than the order of the research questions posed at the outset of the research. Chapter 6 revisits the research questions and provides detailed answers to them to ensure that these have all been addressed in the course of the research. The key findings of this research revolve around a number of key topics. The first topic examines the degree to which a school' outcome-based performance measurement system support the needs of decision makers at both the policy and operational levels. The second topic looks into the convoluted issue of addressing multiple stakeholders' complexity and the ways using foreign consultants can add to the intricacy of the issue. The third topic is related to the much noted concern about lack of resources and its allegedly negative effect on performance measurement systems. The fourth topic is about the role of sustained political support in reform agendas that promote collecting and using data to support transparency and better decision making. The fifth and last topic examines the critical role of well developed internal and external communication strategies to facilitate the introduction and approval of a new measurement culture.

Each research finding will be presented and discussed in turn.

5.2   The Schools' Outcome-Based Performance Measurement System Is Not Providing Decision Makers, at both the Policy and Strategic Level and the Operational Level, with Pertinent Information about the Reform and Its Overall Progress

A review of archival records such as initial reform proposals and publications (Appendix 1 has a list of consulted archival records) reveals that the performance measurement system for the case study organization was designed to be purposefully oriented more toward schools outcome versus schools input and process. Schools outcome measures, such as academic results, are emphasized while schools input and process measures, such as quality of teachers and teaching practices observed inside a classroom, represent a smaller and less important portion of the measurement data (for more details see 4.1.2). As explained in the archival records, outcome measures supposedly capture movement towards the goal of improving student performance; this is the ultimate goal of the education reform according to the designers of the reform.

Academic results of students, a typical example of outcome oriented measures, represent the major component of data collected in the Inclusive Educational Appraisal Scheme (IEAS). These academic results are meant to be key indicators of a school's performance and its progress towards achieving the goals of the reform. The market-driven reform categorizes academic results as the final verdict on the successful progress of its efforts. The idea is to allow schools to choose the best way to operate their own business. Academic results are believed to ultimately serve as objective data that demonstrates to parents and other stakeholder how well schools are really performing. Stakeholders can then decide where to enroll their children according to this data. Competition within the schools' market is expected to improve quality of education in the long term. At the same time, schools are also surveyed for more comprehensive reviews of their performance through the Inclusive School Appraisal Scheme (ISAS). This scheme includes a mix of processes and inputs measures. However, primary data collected for this research

reveals the possibility of negative perceptions towards the credibility of measurement data collected by the ISAS. A number of respondents indicated during interviews and the focus group meeting that the resulting data is generally viewed as less credible due to its self-reported nature.

It follows that the academic results of schools, which constitute the major part of the measurement data, are in essence the main, and seemingly most trusted, source of feedback information that informs different stakeholders about the status of the reform. This finding focuses on decision makers within the case organization and the ways performance measurement data is supporting their decisions at different levels. As mapped on the power-interest grid in chapter 4, these decision makers include an influential and high level member from the policy making body (Leaders and Context setters) as well as six members form the management cadre group (Players).

All seven interviewed stakeholders from the case study organization are directly concerned with the affairs of the reform and have been appointed to safeguard its principles and facilitate its smooth operation; albeit in different capacities and positions. Given the design of the reform, and its assumed dependency on measurement data, it is safe to expect that the measurement data should at least have some level of influence on their policy, strategic, and operational decisions. Indeed, the archival review and data from interviews with stakeholders from the case organization reveal that some key policy decisions were made immediately after the release of academic results (for more details, see 4.1.3). This is not to claim that that the referenced decisions were based on, or consistent with, the measurement data. It only indicates the possibility of a correlation between the two: measurement data and decision making.

This next section catalogs the research findings as far as the different reported reasons affecting the current utilization status of measurement data by decision makers at the case study organization. As the following discussion shows, the findings attribute the current level of utilization to an interesting combination of

reasons that vary from certain characteristics of the current measurement data to a number of emergent premises that are politically and culturally focused. Findings also show that there is a difference in the reasons that are perceived to facilitate or hamper utilization of data to guide decisions at the policy and strategic levels compared to the operational level. Consequently, the findings are presented in two separate sub-sections: one that addresses findings related to utilization of data to guide policy and strategic decisions and another that addresses its utilization to guide operational decisions.

5.2.1 Utilization of measurement data in policy and strategic decisions

There is prevalent agreement among interviewees from the case study organization that data produced by the current measurement system is, in general, not providing them with the information they need to monitor and evaluate the overall progress of the reform. As a result of a number of perceived shortcomings, the majority of interviewees indicate that it is difficult to make informed, high level decisions based on available measurement data. A review of data collected for this research indicates that this view is particularly shared by the interviewee from the governing body and the two directors from the senior management cadre. The governing body is charged with monitoring the reform, developing appropriate policies and setting the strategic direction of the sector it oversees. Although it is the governing body that makes policy and strategic level decisions, the two directors from the senior management cadre have their views about the usefulness of measurement data for policy and strategic decision makers. There is continuous reference by all these public officials, the member of the governing body and the two directors, to the meagerness of measurement data to support policy and strategic decisions because of a number of perceived limitations.

"I would be hard pressed to state that our policy decisions are well informed or influenced by collected data."
(The respondent from the policy maker group)

In view of the fact that the literature highlights supporting better decision making as one of the key values of performance measurement (de Lancer Julnes, 2001), this research's attempt to conduct an inquiry into this matter can contribute to our understanding of the models of measurement data use and the dynamics that affect its utilization. If measurement data in this case study, purposely and by design, is supposed to bear a great responsibility in providing policy makers at the governing body with objective feedback to support better and more informed policy and strategic decisions, then an inquiry into the factors that affect the actual use of data can contribute to our understanding about the dynamics of utilization.

A number of factors that have been shown in the literature to influence utilization of measurement data were incorporated into the interview and focus group questions and explored during the primary data collection stage. These factors include certain characteristics and quality features of measurement data such as clarity, organization, relevance to users, level of analysis, level of details, comparison with past data as well as other features. The following subsections list key factors that were highlighted by interviewees as reasons that affect utilization of measurement data to support policy and strategic decisions.

5.2.1.1 Characteristics of Measurement Data

a) Mix and focus of measurement data

All respondents from both the policy making group and senior management cadre group reference the current mix and focus of measurement data as significantly inadequate to monitor the current status of the reform and help track and monitor its progress and likely future development. As a result, the current mix and focus of measures is perceived to be a factor that hampers the utilization of its data to guide and affect policy and strategic decisions. In particular, the long-term focused measures are increasingly perceived by these respondents as less relevant to policy

and strategic decision makers who are apparently more concerned with monitoring the immediate events that are directly affecting the course of the reform.

It is important to note that this position by the respondents has apparently developed over the years and is a result of the accumulation of working experience with the measurement system. Due to the ambitious and sudden nature of the change, and the apparently unexpected on-the-ground challenges that accompanied it, decision makers <u>are turning their attention</u> to the immediate and shorter term aspects of performance only to find out that there is little or no data that cover these aspects. The current general conviction is that during this early age of the reform, an organization needs to closely follow up what is transpiring on the ground and adjust its policies and strategic positions accordingly. Respondents believe that they can not rely entirely on academic results to evaluate the soundness of their high level decisions.

"At the beginning, we focused on measures that track outcome. But we have since realized that we <u>need to add</u> measures related to what is going on inside the schools. Principals and teachers were not ready for this reform: on- the-ground practices are less than perfect and need to be monitored".
(The respondent from the policy maker group)

This apparent shift in the needs of the policy makers since the inception of the reform calls attention to the original purpose and validation behind the design of the measurement system and raises questions about the importance of revisiting these principles. The rationale for the current performance measurement system was to improve the overall performance of the education system without giving many clues, of itself, to the challenges that exist within any one school. Propper and Wilson (2003) list this model as one of the viable performance measurement systems used in the public sector.

However the entire question about the alignment between the current design of the system and the current needs of policy makers, which have apparently shifted since

inception, is yet to be addressed by those officials in charge of the education sector. Perhaps another question about the robustness of the performance measurement system with respect to changing needs must be posed and examined.

Despite the realization that the governing body needs data from a more 'balanced' mix of measures, both archival records and primary data clearly show that nothing much has happened to change the design of the measurement system or the focus of the measures. The case study organization is still struggling with the overarching question related to what constitutes the 'right mix of measures'. In the reported absence of a review and improvement process that could potentially remedy some of the highlighted shortages revealed by interviewees, measurement data continues to be perceived as fairly irrelevant to policy makers. As a result of the interviewees' negative perception of the currently available feedback information, the utilization of the measurement data in policy and strategic decisions is described by these respondents to be weak, sparse, and not systematic.

"We rarely base our policy decisions on this data." "I don't think our decisions at the policy level are deeply grounded in objective analysis of data."
(The respondent from the policy maker group)

"I was involved several times in policy debates that were influenced by data. However, this doesn't mean that data is systematically used in policy decisions; I don't think this is the case."
(Respondent from the senior management cadre)

b) Analysis of data and comparison with past data

In addition to reasons related to the mix and focus of measurement data, respondents from the case study organization attribute poor utilization of measurement data in policy and strategic decisions to two other factors. Although a review of collected data shows that these factors are referenced less frequently compared with the 'mix and focus' factor, some respondents report (1) **scanty analyses of data** and (2) **poor**

comparison of newly collected data with past data as additional reasons behind weak utilization of measurement data.

It can be assumed that quality analysis of collected data can potentially decipher otherwise raw and abstract data; adding meaning and showing relationship between different variables. On the other hand, comparison with past data demonstrates changes in performance level and makes it easier to track improvement and or deterioration over time. Both reported characteristics are highlighted by the respondent from the policy making body, and one respondent from the management cadre at the case study organization, as reasons that hamper the use of measurement data to inform policy and strategic decisions.

5.2.1.2 Political/cultural factors

Whereas the discussion so far highlights findings related mainly to design issues and the 'technical' characteristics of measurement data, other factors also surface as key reasons that directly affect the use of measurement data in decision making at both policy and strategic levels. Interviewees at the governing body, the operational, and the assessment entities reference two political and cultural factors that they believe are playing a crucial role in impeding the utilization of data to guide policy and strategic decisions. These two factors are discussed in details in the following section.

a) Interest in pleasing the public

Interviewees from the governing body and the senior management cadre draw attention to the role of a prevalent culture that reportedly places great importance on gratifying public sentiments and avoids shocking the system. Such a culture should have been accounted for during the development of the performance measurement system, as reported by some of these interviewees.

"By establishing a system that measures, tracks, and publishes results (and consequently puts individuals in the spotlight), we are taking a serious risk of alienating our society in a culture that values face-saving and has zero tolerance towards embarrassing its citizens. We need to be careful; we are rubbing against the grain of this society".

(The respondent from the policy maker group)

"Assuming that we can simply collect and publish data because the society needs accurate and transparent information is, in my opinion, naive and unrealistic. We need to think carefully about how this society will react and find ways to make sure that publicizing results does not come across as pointing fingers or causing embarrassment to members of this society… This reform needs to follow an incremental and carefully calculated process".

(Respondent from the senior management cadre)

The literature shows that public officials who consider using measurement data experience a number of restraints when they attempt to include results in their decision making process. Hatry *et al*. (2003) reports that reluctance to take decisions and make changes based on outcome information can be caused by the organization's, or it's upper management's, fear of rocking the boat or being perceived as running an ineffective organization. A related idea concerning the "embarrassment effect" of performance measurements has also been discussed in the literature (Mayston, 1985). Measurement data can potentially humiliate public officials if they have to take decisions based on information that reveals inadequate performance or failure to meet promised goals.

However, the concepts highlighted in the literature are 'inward-focused' in the sense that they concentrate on the organization's 'reluctance' and 'fear of embarrassment' associated with using performance measurement data. **This research presents a different focus of embarrassment; one that is interestingly 'outward-focused' and is concerned with fear of embarrassing the society, or members of it, rather than embarrassing the public officials at the organization under consideration.**

Cultural norms, and the prevailing tendency towards pleasing the public, are found to be one of the restraints that face public officials who consider using measurement data when they attempt to include results in their decision making process. Interviewees report that policy decisions have, on more than one occasion, been based on worries about public dissatisfaction and widespread rumors rather than on hard evidence backed by collected data.

Concerns from decision makers about rocking the society's boat show that the public is acting in the capacity of an influential external interest group. Certainly the effect of external interest groups, which include the public, have been reported by organizations that have performance measurement systems as a key factor that influences the success of performance measures (Bowden, 1996; Cannon, 1996). De Lancer Julnes *et al*. (2001) argue that political/cultural factors, such as support from external interest groups, have a preponderance of influence on utilization of data during the implementation stage of performance measurement systems.

One way the support of external interest groups is essential is by allowing using the measurement information even when the results are in breach of a political agenda. For example, a public organization can decide to use controversial measurement information because it realizes that external interest group will support their decision to use the information.

The finding of this research supports the general argument about the role of external interest groups to utilize data during implementation of measurement systems. **Conversely, this research shows that the role of the public, as an external interest group, influences the usage of measurement information regardless of whether or not measurement information is in line with a political agenda.** It can be argued that poor academic results of students might be regarded as further evidence to support the notion about the country's need to reform its education system. Policy makers in the case study certainly have the option to use measurement data, which reflects poor academic performance of the students, to underscore their point about the need for education reform and further push the

reform agenda. However it seems that they are reluctant to use measurement data, even though it can implicitly support their political agenda, in fear of upsetting the public with the immediate bad news about their children's academic performance. **The relationship between support from external interest group to use measurement information and the nature of the political agenda is not established in the findings of this research.**

b) Lack of a measurement culture

The second reported political/cultural reason in this research is the lack of a measurement culture and an overall poor appreciation of the importance of data based decision-making. An analysis of interview scripts reveals that the member of the policy maker group, all interviewed members of the management cadre, and one of the interviewed members of the media point to issues around the 'foreignness' of the concept of organizational measurement, let alone knowing how to use its outcome information.

Lack of knowledge about how measurement data can be used is certainly captured in the literature as one of the reasons that affect the inclusion of results in the decision making process (Hatry *et al*., 2003). However the reason reported in this research goes beyond the technical knowledge of 'how to use data' into a more fundamental question of 'why should we use data'. Apparently, both the case study organization and the society are still new to the culture of measurement.

"The sheer introduction of a modern performance measurement system does not necessarily mean that a shift in beliefs and attitudes towards models of decision making has taken place".
(Respondent from the management cadre group)

"I think that even within our organization people don't quite grasp the importance of measurement; these are all foreign ideas that just found their ways into our organizations".

(Respondent from the management cadre group)

"Ours is not a supportive culture; measurement and data are all foreign concepts that don't resonate well with what people are used to. I would go as far as saying there is anti-measurement attitude. We are not brought up to believe in the value of data and hard evidence. Also, data is linked to accountability in a work culture that never held anyone accountable for his/her work. You end up trying to change so much at once".
(The respondent from the policy maker group)

It can be argued that the combination of the two highlighted political/cultural reasons, a prevailing cultural tendency towards pleasing the society and a lack of a measurement culture, is adding to the challenge of utilizing the measurement data. In a context where cultural considerations towards pleasing the public are so dominant, and the constituents have not yet bought into concepts of measurement, it is not surprising that the idea of basing decisions on 'dry' data, whether or not it is to the liking of the public, is one that is yet to gain acceptance. **The end result is the reported inclination not to utilize readily available measurement data to guide policy and strategic decisions** even if this data (1) supports the political agenda of policy makers, and (2) is readily available to the decision makers through a performance measurement system.

Summary

The perceived inadequacy of current performance measures to provide the governing body with a comprehensive and accurate feedback about the progress of the reform has negatively affected the use of these measures in policy and strategic decisions. The challenges are mainly related to design issues concerning the **mix and focus** of the measures but also related to potentially human capacity issues reflected in the absence of strong **data analysis** and poor **comparison with past data**. These perceived 'weaknesses' seem to render data produced by the performance measurement system as largely irrelevant to policy maker and not in line with their

specific needs. Other key reported reasons that are hampering measurement data utilization include **poor cultural appreciation of data based decision making**, and cultural tendencies to make decisions that are geared towards **pleasing the public** (regardless of what the data reveals). Figure 5.1 summarizes the different key factors that reportedly affect utilization of performance measurement data in policy and strategic decisions.

Figure 5.1: Factors Affecting Utilization of Measurement Data in Policy and Strategic Decisions
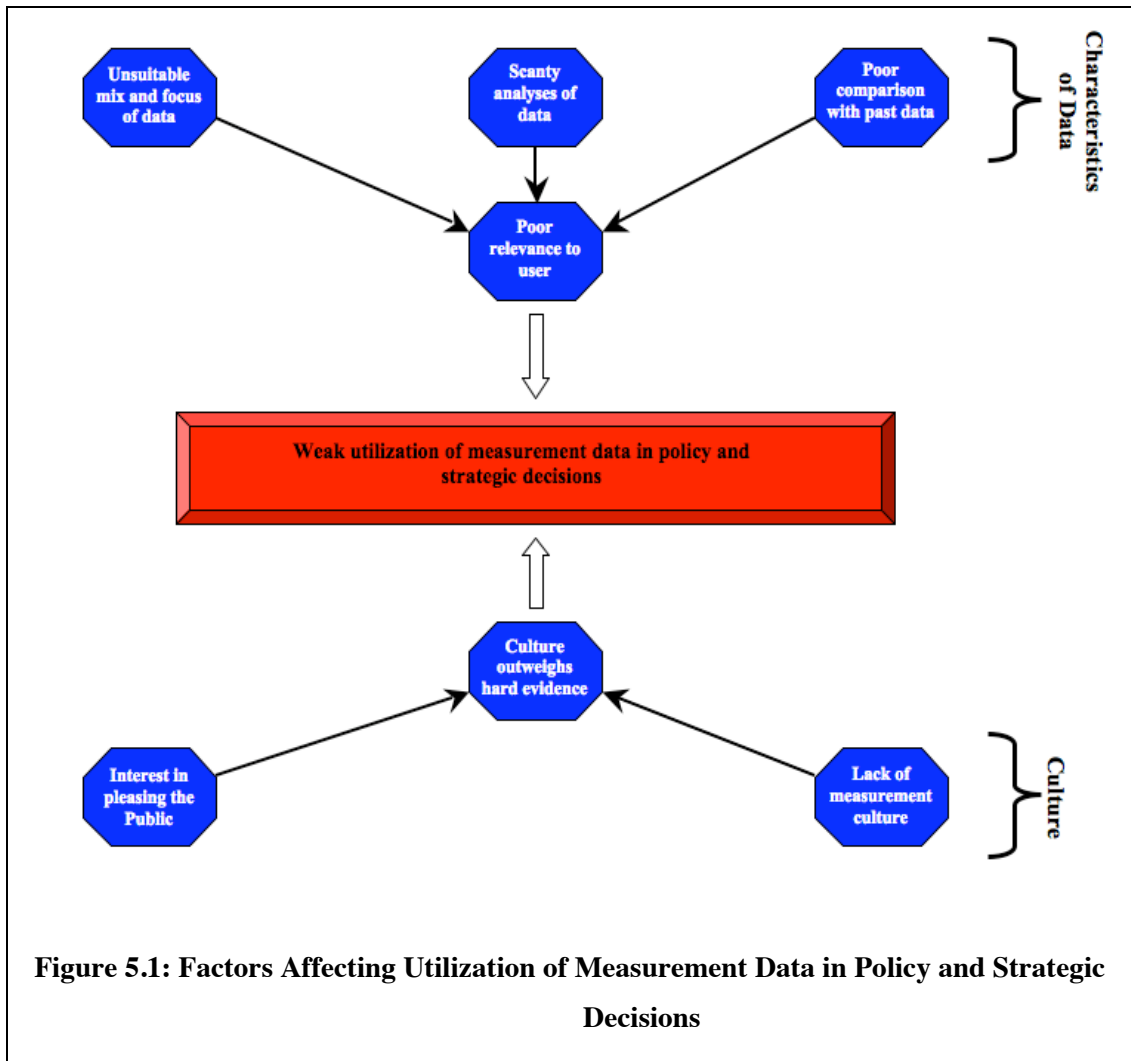
5.2.2 Utilization of measurement data in operational decisions

Archival records show that operational decisions related to the new reform schools that operate under the jurisdiction of the case study are made routinely by the Operational Entity and occasionally by the Assessment Entity. A key part of the Operational Entity's responsibilities is to monitor the ways in which schools implement the approved policies and strategies of the reform and make appropriate decisions accordingly. Operational decisions can be defined as decisions taken by public officials at the case study organization regarding the daily and short-term practices and dealings of the new reform schools. These mainly include staffing, building, maintenance, financial related matters, student affairs, teaching pedagogy, and curriculum development.

The operational decisions supposedly stem from, and help reinforce, policies and strategies adopted by the governing body. Moreover, the Operational Entity is entitled to take punitive actions against a school, including revoking its license to operate, if the entity decides that the school's day to day operations are not in line with the reform's guidelines, policies and strategic direction. Making such punitive decisions is a grave responsibility and all interviewees from the Operational Entity feel that such decisions should ideally be based on certain and indisputable information about operational practices.

All interviewees from the **senior cadre management group** at the Operational Entity believe that data produced by the current measurement system is not providing them with the kind of information they need to monitor and evaluate the day to day progress of the reform. The performance measurement system does not address the information needs of the Operational Entity regarding short term issues of teaching practices, curriculum implementation, student development, and other day to day practices. This is negatively influencing the extent to which measurement data is being utilized to guide decision making at the Operational Entity even though these interviewees unanimously report that they wish to base their decisions on

objective data. This widespread perception of measurement data' inadequacy to support decision making at the operational level is attributed to a number of reasons.

5.2.2.1 Characteristics of measurement data

a) Mix and focus of measurement data

One of the most often cited reasons behind poor utilization is the outcome-focused nature of the measures. The mix and focus characteristics are apparently not quite analogous with the requirements of public officials charged with overseeing the daily actions and processes of an ever expanding number of reform schools. According to interviewees from the Operational Entity, monitoring implementation reportedly requires a different focus of data collection; measures that cover input and processes are believed to be more relevant to address operational issues especially in high stake situations where the question about the future of a school is posed.

"Complex problems surface daily from implementation realities; these can't wait unresolved until we find out what students score"… "data might be good for others but not to us" ..."the consequences of reaching a decision that a school doesn't function well are huge and can't be based on a criteria that we find lacking."
(Different Respondents from the senior management cadre group)

The literature indicates that the question about how to evaluate the efficiency and effectiveness of policy implementation, and whether the phenomenal increase of output measurement in the public sector is the best answer, can raise a contentious debate (Van Thiel and Leeuw, 2002). At the same time, there are studies that reveal a shift in the form of performance measures used in the public sector from collection of data on a narrow range of dimensions (such as outcomes) towards development of more streamlined and balanced packages of indicators (Mannion and Goddard, 2000). This shift in focus suggests a trend that is more holistic and better resonates

with the recent growth in balanced frameworks as discussed in the literature review chapter.

A review of the use of performance measurement in the education sector across the world actually reveals a more flexible approach that is not necessarily wedded to an 'outcome-only' mode of measurement. The UK, for example, employs a full spectrum of performance measures that cover processes, outcomes, and measures derived from administrative data. The UK's Office for Standards in Education (OFSTED) focuses on measures derived from an in-depth evaluation of a school's processes and outcomes. In addition, the UK uses truancy rates and limited measures of national examination pass rates are published for UK schools (Propper and Wilson, 2003). As a policy implementing agency, the Operational Entity is grappling with a performance measurement system that is designed to improve the overall, long-term, performance of an education system without providing many indications about the problems within any given school.

In addition to concerns about the implications of failing to monitor the different aspects of policy implementation, two out of the three interviewed public officials at the Operational Entity are apprehensive about drawing conclusions and making judgments based on academic scores alone. These concerns extend beyond the need for a performance measurement system with a holistic appreciation of different factors within a school to the effect of factors external to schools' practices.

"They expect 'us' to revoke a school's license based only on how well students are doing. This is so unfair on many counts. We would be encouraging and supporting schools that are lucky enough to be open in rich neighborhoods, where children already have knowledge of a 'foreign' language and are more exposed to modern education standards, while penalizing schools that might actually be working harder but have poorer students from uneducated families. I have an ethical problem with this."
(Respondent from the Senior Management Cadre/Operational Entity)

"Most of the parents can't even read and write in these suburbs and their children are put in an immediate disadvantage when they are tested against the same standards used for children of affluent families. We seriously need to think about our monitoring system. Maybe an incremental approach is called for."
(Respondent from the Senior Management Cadre/Operational Entity)

Indeed issues about the relationship between students' academic achievement and factors external to schools' practices and procedures are widely discussed in the literature (The Coleman Report, 1966; EdWeek, 2006). The effects of socio economic factors, for example, are believed to be key factors that are directly influencing the academic outcomes of the reform schools as reported by the two public officials from the senior management cadre group. Schools that operate in certain affluent neighborhoods are more likely to achieve better academic scores on newly introduced standardized tests. These public officials believe that this is not a reflection of a good management approach or better pedagogy practices for example; rather, good academic results are probably related to students' comparatively better base knowledge of the foreign language that is now used to teach most of the subjects. The use of raw outcome scores that are level based, such as the one used in the case study organization, has been criticized in the literature as subject to bias as these do not consider factors outside the school's control (Kane and Staiger, 2002; Propper and Wilson, 2003).

By the same token many respondents in the focus group express reservations about penalizing schools that operate in rural areas, where the likelihood of poor academic achievements are high. During the focus group discussion, some of the schools' administrators and teachers share their reluctance to operate and work in rural neighborhood that are reported to be predominantly inhabited by low income families with the vast majority of the parents being poorly educated. Using level-based academic outcomes of students as the fundamental marker to guide operational decisions is frowned upon by interviewed public officials charged with operational monitoring as well as by schools' administrators and teachers.

b) Data release frequency and tardiness

If substance-related, and at times ethical, concerns about the focus and mix of measurement data are negatively influencing the choice of public officials to utilize the data in their operational decision making process, other practical factors surface as additional reasons that are adding to the challenges of utilization. Issues related to the frequency of data release, as well as chronic delays in releasing the data, are cited by all respondents from the Operational Entity as actual stumbling blocks that render the data not viable.

There is a reported high level of dissatisfaction with the current frequency of data release. The existing pattern of data release is part of the system design and follows set data collection timetables that take place on annual basis (as discussed in 4.1.2). Interestingly, the data release schedule seems to be out of synch with the needs of public officials charged with taking a number of key operational decisions based on the findings of this collected data. The current annual release of the data is scheduled to take place at the very end of the academic year. This is perceived by the interviewed officials as being too late in the process to be of any value to them. An end of academic year data release does not allow public officials at the Operational Entity time to interact with and guide the schools before they close down for their long summer breaks. In the absence of feedback and reaction from the schools, public officials are unable to take guided and well-informed financial, staffing, and other operational decisions related to the schools. The schedule also makes it impossible for the schools to have time to analyze the findings and take corrective measures through their academic and action plans for the next academic year.

"We face quite a challenging implementation reality; this is a very progressive and challenging reform. We can't put off taking decisions until data is released once a year."
(Respondent from the case organization)

"They (operational entity) expect us to provide them with our plans to address our weak areas as revealed by the data. Otherwise they can take a number of penalizing measures against us. However, and under the best case scenario, we don't get the results of the collected data until the beginning of the next school year. The arrangement simply doesn't work"

(School Administrator Respondent from the focus group)

In addition to perceived flaws in the data release schedule, a recurring pattern of tardiness in the release of the data, since inception of the reform, is adding to the problem. All interviewed public officials at the Operational Entity believe that tardiness is causing data to be outdated and hence useless to any meaningful utilization such as guiding decision making. Respondents from the Assessment Entity offer a number of justifications to explain the recurring delay but nonetheless admit to the continuing existence of the problem, despite a reported gradual improvement over time.

"It's always too late to use the data by the time we get it. How can we take a decision about a school when we receive data about their performance almost an academic-year later?"

(Respondent from the Senior Management Cadre/ Operational Entity)

Given the strong link between the design of the performance measurement system and existing patterns of its data release, the reported dissatisfaction with frequency is perhaps more difficult to address compared with tardiness. As capacity building and experience accumulates, tardiness is improving over time. However the issue around patterns of data release is likely to require a more serious discussion about the entire design and set up of the performance measurement system.

c) Format, analysis, and comparison with past data

In addition to focus and mix, frequency, and punctuality, the format used to disseminate collected data is reported as another reason that affects data utilization in

operational decisions. **Open files formats**, reportedly unorganized and loaded with too much details, are used to circulate collected data within the case study organization. This format is cited by concerned respondents as an obstacle that makes reading and understanding the files difficult and time consuming. Moreover, data included in these files lack **good analysis** and offer **poor comparison** with past data as reported by the majority of interviewees from the management cadre group.

The perceived negative attributes of the measurement data add to the reasons that are hampering its utilization to guide operational decisions. Public officials report that they don't have the time, training, and right set of skills to sort through masses of data, analyze and infer meaning from raw data, and place results in its sequential order to draw accurate conclusions about progress and change over time. This also suggests the need to slice data according to the needs of the different recipients groups and use a more structured format to present the data to its users.

Summary

Concerned decision makers at the Operational Entity perceive the current measurement data as lacking and incapable of providing them with the kind of data they believe they need to carry out their mandates. The main concern is related to the **mix and focus** of collected data. This perception has negatively affected the use of data to shape and influence their operational decisions. Other factors related to **frequency and tardiness of data release**, **format of files**, **level of analysis**, and **comparison with past data** are shown to also play a role in affecting the degree to which data is being utilized in operational decision making. Unlike findings about policy and strategic decisions, political and cultural factors are not referenced as reasons that influence the choice of interviewed public officials to utilize measurement data in their operational decisions. Figure 5.2 summarizes factors reported to affect utilization of performance measurement data in operational decisions.

This finding of the research does not support the findings of de Lancer Julnes *et al.* (2001) which highlight factors addressed by political and cultural considerations as having a preponderance of influence on utilization of performance measures during the implementation stage. A respondent from the senior management cadre articulates that since their mandate is to "merely implement policies set by higher authorities" they don't need to worry about aligning their operational decisions with the public's outlook. This position seems to be the general conviction of most respondents, including those who are otherwise concerned about the media's negative position towards the reform. Interviewed public officials are apparently more interested in ways to improve utilization of current measurement data to guide their operational decision than in catering to the public's sentiments and attitudes.

In addition, **this finding does not support the findings of Bourne (2005) regarding the reasons that influence management commitment to performance measurement systems over time.** The case study' management commitment to the measurement system, in both absolute and comparative terms as discussed in chapter 2, is seemingly existent and unwavering. Although it is logical to expect that at least some shift of priorities has taken place since the inception of the reform, attention paid to the implementation of the performance measurement system by management does not seem to have been adversely affected. Reasons related to impatience and resistance are not reported in the collected data and do not seem to apply to the case study organization.

Figure 5.2: Factors Affecting Utilization of Measurement Data in Operational Decisions
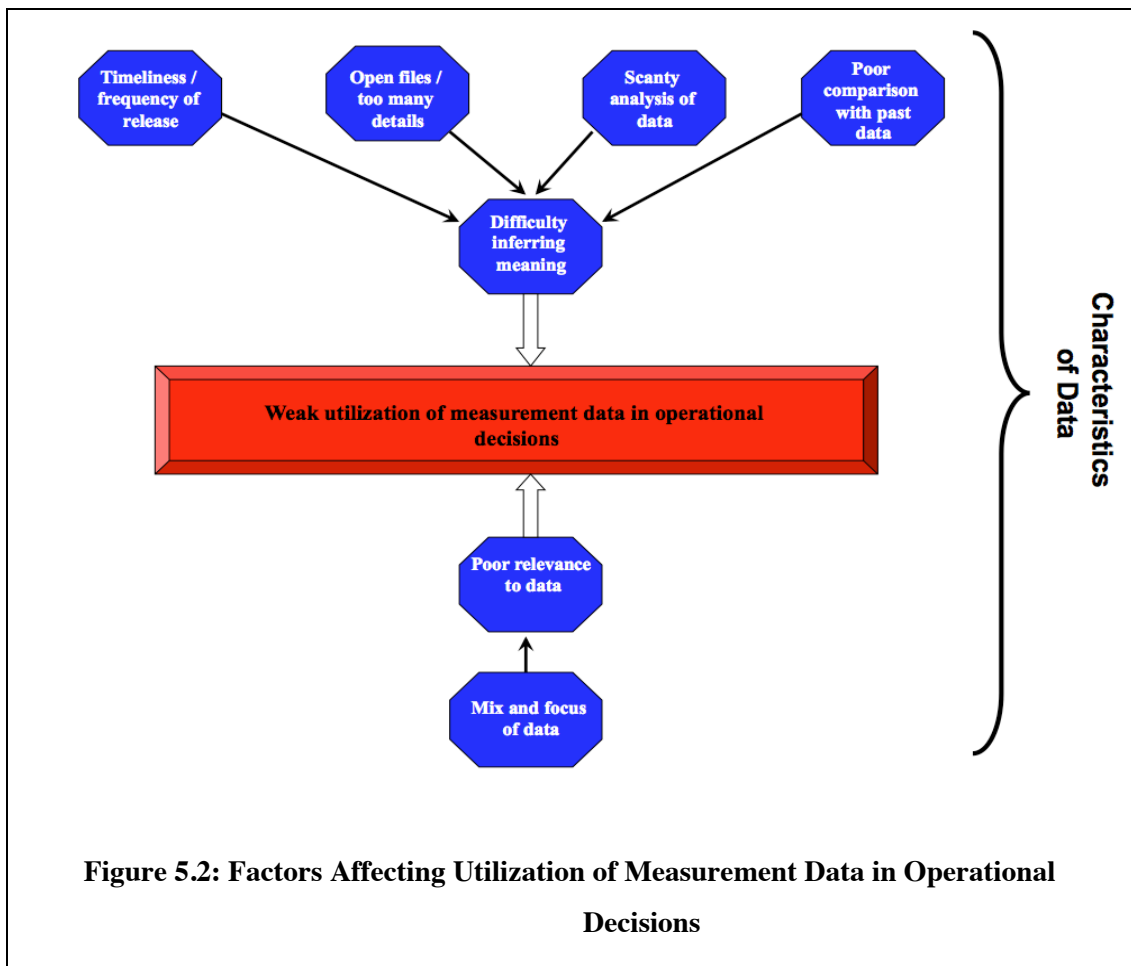
5.3   An Imported and Modern Performance Measurement System Has Not been Shown to Create Satisfaction amongst Local Stakeholders

During the data collection stage, respondents were thoroughly questioned and probed with different scenarios to reveal their perception and satisfaction level with all aspects of the performance measurement system: its processes, measures, formats of reports, timeliness, and relevance among many other factors that are referenced in the literature. The researcher was looking to achieve a deep understanding of how measurement data was perceived by different stakeholders before moving to 'how' and 'why' data was later on utilized in any given way.

While a number of issues are highlighted as pertinent to the discussion, two interrelated issues are frequently referenced and discussed with considerable fervor. The first issue circles around the foreign origins of the measurement system and the generally negative disposition toward foreign affiliations in the education sector. The collected data also reveals a second issue regarding a widespread perception among stakeholders that the foreign consultants neglected, or even failed, to involve different local stakeholders during the initial stage of designing the measurement system. This has resulted, in their views, in a system that seems to be incapable of tracking a number of issues that are critical to stakeholders and address their specific needs and concern.

5.3.1   Foreign origins of the performance measurement system

Haylett's (1964) definition of consultation as a professional activity in which a specialist attempts to help a less knowledgeable consultee solve a problem applies to the nature of the relationship between the case study and the foreign organization that designed the performance measurement system. Caplan (1964, 1970) restricted consultation to an interaction between two professional parties: the consultant or specialist and the consultee who has asked for help in regard to a work problem. This work problem, according to Caplan, includes the planning or implementation of a program designed to serve the needs of the consultee; another definition that fittingly describes the nature of the relationship between the case study and the foreign organization that designed the measurement system.

The fact that the performance measurement system was designed by foreign consultants seems to be well known to all respondents who were interviewed in this research. At the same time this fact seems to be, in some ways, working against adopting and accepting it by some stakeholders. Responses swing between claimed capacity to develop similar systems without the need for foreign help and fears of ill agendas to steer future generations toward relinquishing their heritage and abandoning their national culture.

The results of the media coverage review conducted by the researcher reinforces this finding and reveals the existence of local concerns about foreign plans to manipulate the new generation and influence their traditions and religious roots. The following selection of excerpts reflects the trend in thinking among different groups of stakeholders regarding the issue of foreign affiliations:

"The entire system is imposed on our culture and not sensitive to our needs…we need to look back and think how this system was developed and by whom". "Were there any local people involved during the planning stages? …our country is worried about losing its national identity and heritage, we feel that we can develop our own systems in a way that suits our society and needs"

(Different Respondents from the media group)

"There is a lot of suspicion in the community as to why our education system was set up in a way that focuses so much on English language…this will be at the expense of our language and culture…If our children are weak in their own language they will stop reading books in that language, they will stop reading our holy book. This is an intended recipe for losing our identity and heritage"

(Respondent from the case organization)

At an even more skeptical end of the spectrum of comments offered, a respondent from the school administrators group stated that she believed that "*foreign consultants kept us out of the picture for a reason; everything was cooked up away from the people concerned and suddenly it was introduced to the country*". Kaffman (1961) citing the early comprehensive work of Caplan (1960) on the topic of foreign consultants, brings attention to the patterns of interactions between technical consultants who operate in foreign countries as consultants and consultees. In particular Caplan points that the types and patterns of interactions between the two groups are likely to be shaped by the inevitable differences in the specific conditions in each country. Kaffman (1961) argues that the success or failure of plans developed by consultants does not depend entirely on their skills and technical abilities; rather it can sometimes be adversely affected by a number of emotional factors that inherently exist in the style and pattern of interactions between

consultants and consultees. In addition to the obvious challenges that are likely to be posed by language differences, the collected data in this research reveals a more deeply rooted issue related to sentiments towards the West and, as a result, Western consultants; which is where the foreign consultants for the case study are from.

"It was designed by foreigners … we hear about what the West has in mind for our children; their plans are now built into our system, what we focus on and, as a result, what data we collect and publish"
(Respondent from the school administrators group)

Further examination of the collected and archival data, especially of the reform background and beginnings, raise a question regarding whether or not the top-down nature of the reform in the case study organization might have contributed to the prevalent unease towards foreign affiliations. It may be that adopting a more participatory approach, or maybe even a more transparent and efficient process to transmit information about the development of the reform and its measurement system, and the underlying principles that guided its design, would have alleviated some of the suspicion regarding original intentions. Indeed data review reveals concerns about both internal and external reform-related communications; a point that is separately discussed later on in this chapter. Regardless of other factors that can be playing a role in agitating negative local sentiments against foreign consultants, the collected data clearly reveals serious apprehension among stakeholders regarding foreign affiliations that merits listing under a separate finding in this research.

If the ideas and plans that are presented by foreign consultants operating in foreign countries are not necessarily judged solely on their objective values, and if attitudes emanating from psychological factors in the background of consultees, as well as of consultants, do indeed play such a major role, then this might have added another layer of intricacy to the entire issue of accepting and, as a result, utilizing this 'foreign' performance measurement system.

The apparent complex and tense dynamics of the relationship between foreign consultants and consultees can be traced in the script of some of the interviews and focus group. This research did not look deeply into understanding the underlying reasons behind the nature of this relationship mainly because the entire point about the foreign roots of the measurement system surfaced as an emergent theme out of the collected data. Moreover, the researcher noted a paucity of research in the subject about the relationship and dynamics between foreign consultants and local managers, especially in developing countries. Given the focus and design of this research, and the scarcity of published work under this particular subject, it is difficult to make any further inferences from the collected data despite the frequent referencing to this point during the interviews and focus group.

5.3.2   Addressing multiple stakeholder complexity in PMS design

Not only is the performance measurement system viewed with suspicions, and at times resentment, at least partially because of its foreign roots as reported by some of the respondents, stakeholders are equally unimpressed by an imported design that is not quite sensitive and responsive to their data needs. Different stakeholders' groups seem to believe that the measurement system, as elaborate, modern, and state of the art as some of them think it is, neglects performance aspects that are important to them and fails to collect data that is at the heart of their key needs and priorities.

At the same time, there is a high degree of variation amongst respondents as far as what constitutes 'important data needs'. This finding reveals the quite diverse data needs of the case study's key stakeholders and the apparent mismatch between what **they** think of as important information and **what kind of information is provided** by the performance measurement system.

The topic of multiple stakeholders seems to be at the very heart of any discussion about public sector management. In fact, it is often argued that the issue of multiple stakeholders is unique to the public sector and poses a number of challenges that must be addressed before rushing into adopting many of the private management

techniques that are often promoted in the philosophy of new public management (Boyne *et al.*, 2004; McAdam *et al.*, 2005).

In addition, there is seemingly an agreement in the literature that multiple stakeholders can have multiple and, at times, conflicting agendas and needs (Rowley, 1997; Dixit, 2002; Propper and Wilson, 2003; Bryson, 2004; McAdam *et al.*, 2005). Different stakeholders have different requirements and expectations from public organizations. Bryson (1995) highlights the importance of paying attention to stakeholders because the "success" of any public organization depends on satisfying key stakeholders according to "their own definition of what is valuable".

Performance measurement systems can be regarded as a management instrument which guides the current and future direction of public organizations. While there are many factors involved in performance measurement systems, Bendheim and Graves (1998) argue that all of these factors are influenced by the overarching effect of multiple stakeholders. This is further reinforced by warnings that failure to attend carefully to the interest and information needs of stakeholders "can easily lead to disaster" and that "the complexity and multiplicity of stakeholders must be accounted for in any performance measurement system" (Bryson, 2004 citing the work of Nutt, 2002).

Wisniewski and Stewart (2004) highlight the importance of having differing "stakeholder windows" on public sector performance measurement where "a one size fits all" approach is not likely to be successful. The critical need to address the complexity of multiple stakeholders in the public sector must be integrated into the design stages of any performance measurement system. Bryson (2004) strongly cautions against excluding stakeholders during the development of performance measurement systems.

The fact that stakeholders of the case study organization did not partake in the development of the measurement system could well be contributing to their negative attitudes toward the current system. Studies show that mandating performance

measurement alone does not work and that the development of measures requires an approach whereby stakeholders and performance measurement experts, or consultants, work together at developing measures that are meaningful to stakeholders and for which they will eventually be held accountable (de Lancer Julnes, 2001; de Lancer Julnes, 2006).

"As a director, I was never consulted about what the new measurement system should be like or what areas it must focus on. We just inherited the system and were instructed to use it"
(Respondent from the case organization)

All interviewed members of the **senior management cadre** at the case study share some level of discontent about not being involved or conferred with during the system design and measures development stages. However given that this young reform was started from scratch, it is not quite surprising that current employees in the case organization were not involved. Many of them were simply not working at the organization at the time and it was virtually impossible to identify members of the future management staff. Nevertheless, this fact does not seem to alleviate the general theme of disappointment as a result of the exclusion of this key stakeholder group from the initial discussion about the measurement system.

Perhaps the situation is worsened by the nonexistence of a review process, or any similar effort, that would otherwise, and with time, incorporate stakeholders' feedback into the measurement system. Indeed, further contemplation of the collected data points to the possibility that the static nature of the measurement system might be complicating the problem. Stakeholders, in particular within the case study but also external ones, continue to be 'forced' to use, or receive information from, a measurement system that they had no say in its design in the past and have no say in improving it for the future.

In an attempt to add order to the collected data, and to better understand the wide range of data needs that are identified by different stakeholders in this research, the

following sub-sections group these findings into three main categories of data requirements. Figure 5.3 represents a summary of these categories.

Figure 5.3: Missing Measurement Data Needs of Stakeholders
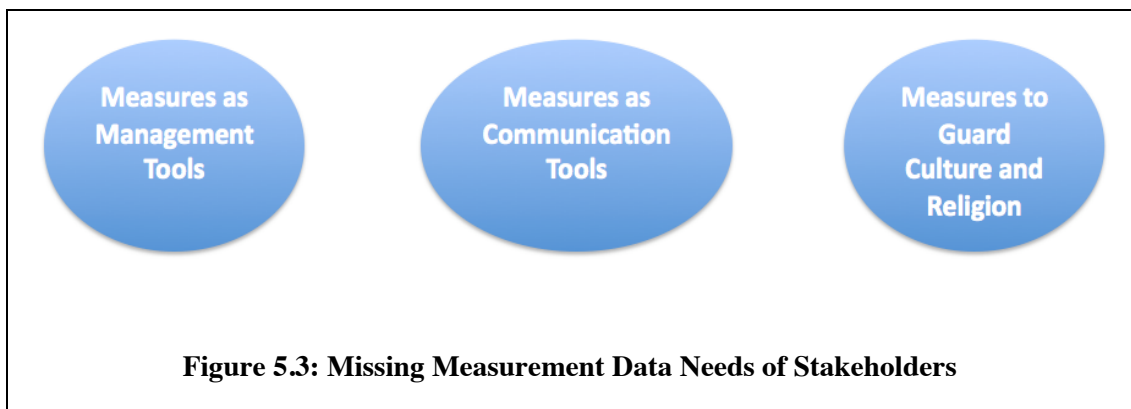
5.3.2.1.   Stakeholders' need for measures as management tools

Most respondents from **the case study organization** recognize a need for measurement data to help them monitor and manage the internal performance of their organization and its different units. The use of performance measurement information as a management tool is recognized in the literature as one of the various possible ways in which these systems may be used (Burgess *et al*., 2002, Propper and Wilson, 2003). Because such internal data is reportedly not collected in this case study, decisions related to the management of the case organization are not benefiting from or guided by data collected by the current performance measurement system. Interview data also reveals that the same respondents share a general lack of clarity about organizational goals and strategic direction. There are prevailing concerns among this group of respondents that the organization is flying in the dark with no management tools to guide its operation. The archival review indeed reveals that, despite the wide range of data that is currently being collected by the measurement system, there are no measures that track or collect data related to the internal performance of the case organization itself.

"The system doesn't serve the internal needs of my department…. it doesn't give me any information about how well we are doing, whether we are on the right track of performance"…"I'd like the system to include internal measures…which can help me manage daily work but also strategically plan for the future"…"this is a phenomena across the organization, no one knows exactly what they are doing or where they are heading; we are flying in the dark. We don't even agree on what this organization's main goals are"
(Different Respondents from the case organization)

The literature points out to the relationship between **the existence of, and consensus on, strategic direction and program goals** in an organization and the **likelihood of adoption and implementation of performance measures** (Attkisson *et al*., 1978; de Lancer Julnes, 2001). This relationship intrinsically assumes that at least some of the performance measures an organization aspires to adopt and implement are linked to strategic and program performance aspects. Perhaps this finding in the literature is based on the assumption that identifying internal organizational performance measures is a given; a key component of developing any measurement system. This is not quite a surprise given that the literature does offer a number of frameworks to guide the development of performance measurement systems; all of which address internal organizational processes and goals one way or another (see 2.5). In fact some of these frameworks are presented as viable strategic planning tools that can close the loop between performance measures and the strategic direction of an organization (for an example, see Balanced Scorecard under 2.5.1).

However, the situation in the case study is different in that this type of performance measures is entirely missing from the overall measurement system. This creates a unique situation where some of the basic assumptions in the literature might not necessarily apply to this case study. This also brings to the discussion a question that surfaced separately during the archival record review: **why is there no evident utilization of any of the many performance measurement frameworks available in the literature?** Internal processes and other aspects of organizational performance

are clearly important enough and are seriously considered and included in all performance measurement frameworks.

The other noteworthy finding is the kind of relationship the respondents envision between 'internal organizational measures' and 'strategic direction and goal'. These respondents report that they find it difficult to set their future direction and draw their plans due to absence of "*internal checks and balances to help 'us' read our current accomplishment picture…we run from one project to another without overall guidance from objective measurement data that would otherwise make our current and future map for action much clearer*".

There seems to be a perception that a causal relationship exists between the two factors: identified internal organizational measures can help these respondents clarify the organization's strategic direction and program goals. In other words, the development of internal organizational measures should take place first. This would then lead to identifying the strategic priorities and setting the program goals of the organization.

This spread belief between this group of respondents, all of whom are from the case study organization, is quite distinct from the general conviction in the literature which typically points to a process that works in the opposite direction: identification of strategic priorities and goals precede the development of performance measures (see 2.5 for a review of the steps inherent in some of the frameworks discussed in the literature review chapter). While it is unclear why this is the case, it is quite plausible that the position of these public officials is an indicator of their high expectations of their organization's performance measurement system. They seem to view this system as a tool that is capable not only of addressing their internal management challenges but also helping them identify their strategic direction and priorities. Without relevant organizational performance measures public officials at the case organization find it challenging to set their strategic direction and get internal consensus on the true goals of their programs.

166

5.3.2.2. Stakeholders' need for measures as communication tools

The need for measures that monitor internal performance extends beyond providing the case study' officials with a management tool that can help them track progress and monitor the organization's internal affairs, as important as these functions are regarded by the concerned respondents. There is apparently a need for internal data to satisfy public scrutiny, inform the public about the case organization's internal accomplishments, and keep the public abreast of progress made toward achieving the overarching goals of the reform.

This particular type of data need is identified by **the two respondents from the media** and **the two respondents from the senior management cadre** at the case study organization. The perceived inadequacy of the measurement system to provide information about the accomplishments of the case study and highlight the comparative advantages of the new reform system is expressed in the following excerpt:

"We achieved a number of critical improvements in the country's education system. But the system failed to capture our internal success stories. No one outside our organization is even aware of our performance and we feel neglected and unappreciated".
(Respondent from the case organization)

In this context, the absence of measures to monitor internal organizational performance is particularly worrisome because the case organization is a public entity that was established outside the organizational and legal norms of the rest of the public organizations in the country (See 4.1.1). It was bestowed with many privileges which, as a result, probably turned it into a magnet for scrutiny by skeptics of the reform. Data that support the soundness of organizational decisions, showcase progress toward set goals, and at times defend the legitimacy and integrity of sometimes unorthodox public processes are in great demand to counteract attacks on a reform largely symbolized by the case organization. Interestingly the need for this

kind of information is not unique to the public officials at the case organization; other stakeholders in the society are equally interested in the internal affairs of the case organization:

"We are very keen on understanding how this new public organization is running its business; the entire country is curious about this matter. To be honest, I haven't seen any data that can help me personally understand how this organization is operating."
(Respondent from the Media)

5.3.2.3. Stakeholders' need for measures to guard cultural and religious interests

In addition to identified needs for measurement data that can be used as management and communication tools, **most participants in the focus group** and the **two interviewees** from the **media group** articulated yet another need for measurement data to address an issue of high priority in the local society. The issue focuses on the importance of preserving local heritage, tradition, and religious beliefs. This calls for the collection of data that measures and tracks effort and progress made toward preserving these values that are seemingly critical, at least to some group of stakeholders. Respondents find it particularly important that a performance measurement system for the education sector provides them with this kind of information.

"The system needs to be tailored to the cultural needs of our society. This is its biggest weakness…. We need to know how our children are doing as far as our religious and cultural beliefs are concerned…. There is missing information that the society needs to make informed decisions about a school; it's not only about English language, grades and academic achievements".
(Respondent from the media)

"I feel some information is missing about how well the school meets my family's needs, the information comes across as too generic". "I would like to learn more about the background of my children' teachers; their beliefs and track records, how well their ideas fit with our culture…many teachers are non-locals and we don't really know their background or views on different issues that are important to us".
(Different Respondents from the parent group)

While most respondents to this research view the performance measurement system as insensitive to their information needs, this group that identifies the need for information to guard cultural and religious interests seems to be the most troubled with the consequences of this 'information gap'. One respondent thinks that '*it's just a matter of time before the country gets rid of this measurement system that is of no use to the society and probably poisoning the minds of the citizens by diverting their attention to meaningless issues*'.

It is worth noting here that concerns about foreign affiliations discussed under 5.3.1 above are strongly present during the discussion with this group of respondents. A careful review of the collected data shows that **all respondents** who covey their need for measures to safeguard culture and religion also make reference to the issue of foreign connections, and or origins, of the measurement system. This research finding clearly indicates that at least some stakeholders perceive some kind of r**elationship between the two points: foreign education fingerprints, and, missing measurement data to monitor and track adherence to culture and religion**. There is an implied allegation of purposeful scheming that resulted in a measurement system that pays no heed to important, and sensitive, cultural and religious information needs. The researcher is not able to identify relevant work in the literature body that can guide the discussion about this seemingly important topic.

"It is nice to have strong math skills but this is something our students can pick up later on in some Western college. However if our students are not well educated about their religion and culture, there is nothing any college can do to fix this

shortage: they are lost forever. This is the kind of information we want to know, this is what counts"

(Respondent from the teacher group)

Summary

Successful public reforms are typically those that have the strong support of key stakeholders (Cabanero-Verzosa and Garcia, 2009). It is conceivably difficult for this research to arrive at a firm conclusion as to which is playing a more influential role in the stakeholders' negative perception about the system and its ability to address their needs: the fact that it was designed by foreign consultants or the apparent negligence in addressing the complex issue of multiple stakeholders during the system's design stages. However, both factors are seemingly playing important, and at times interrelated, roles.

It is also quite evident that what seems to be an over-simplified approach to stakeholders during the development of the measurement system is playing a role in the ultimate poor utilization of performance measurement data. The supposition of a unitary approach which focuses mainly on the schools system, namely on its outcomes, has led to the negligence of the needs of other stakeholders, including those within the organization itself. This could be playing a role in the stakeholders' general disapproval of, and consequently poor use of, the measures that currently constitute the performance system.

In addition, it seems that failure to address the differing agendas and requirements of the multiple stakeholders during the design and development stages of the performance measurement system has only pushed problems of disagreements and incongruities to the current policy implementation stage. The expectations of certain groups of stakeholders are high, or at least not in line with what the system was originally set up to achieve. This finding is consistent with the literature and is at the core of the uniqueness of stakeholders at the public sector.

To address challenges associated with the design of performance measurement systems, the literature offers a number of frameworks that probably could have been used to guide the development of the new performance measurement system at the case study. For example, Moullin (2002) identifies the performance prism model that was developed by Neely as a framework that can address the voices of multiple stakeholders within public organizations. There are also a number of methods that can help facilitate the analysis of multiple stakeholders prior and during the development of performance measurement systems (Ackerman and Eden, 2003; Bryson 2004).

This discussion then begs the question about why there is no evidence that any of these approaches and frameworks were utilized during the design of the current system. Perhaps the foreign consultant realized that shortcomings with participatory approaches that have been suggested in the literature are yet to be addressed by further research. De Lancer Julnes (2001) highlights a number of shortcomings with existing participatory approaches. The notion of having to address several, and sometimes conflicting, stakeholders' requirements during the design of a performance measurement system has not yet been matched with research that clearly puts forward ways and means to systematically identify and analyze stakeholders and manages their relationships (Bryson, 2004; McAdam and Henderson, 2004). It is also plausible that given the top-down nature of the reform, the foreign consultant may have not seen a need to pursue the thorny mission of identifying and addressing the needs of multiple stakeholders.

It can also be argued that the focus of the performance measurement system in the case study organization is purposely oriented towards the schools' long term outcomes. While this design can probably serve matters of parental choice and improvement of quality of schooling in the long run, it resulted in earnest limitations in areas that can be considered critical to the success of such a major reform. For example, enabling the organization charged with spearheading the reform with tools to better manage its internal affairs, and empowering it with means to demonstrate its

capacity to run the reform, can arguably be regarded as two goals that are likely as important as facilitating and supporting school choice.

5.4   Abundant Financial Resources and Modern IT Systems Don't Compensate for Lack of Human Capacity, Organizational and Legal Setups

This research examines a number of rational constructs that are identified in the literature as factors that affect the utilization of data produced by performance measurement systems. Attempts to understand utilization of data, as a step included in the process of performance measurement systems' implementation, called for drawing on findings from other fields such as evaluation programs' implementation (de Lancer Julnes, 2001). Committing resources, financial and human, are among the key rational factors identified in practice and literature to have a strong impact on the implementation of program evaluation (Holzer and Halachmi, 1996) and utilization of measurement data (de Lancer Julnes and Holzer, 2001; de Lancer Julnes, 2009).

5.4.1. Commitment of financial and human resources

Interviews and the focus group indicate **an almost unanimous agreement regarding the abundance of financial capital** that supports the reform in general and the measurement system in particular. In fact, collected data points to a profusion of modern buildings, equipment, and IT systems. As discussed in chapter 2, the readiness of an organization for performance measurement systems includes logistical readiness and availability of IT systems to facilitate the collection, tracking and analysis of data (de Lancer Julnes, 2001). The absence of adequate equipment and IT systems to facilitate this effort can negatively affect an organization's readiness to implement and utilize performance measurement systems.

Despite a reported tendency to increasingly employ more restrictive financial policies, mainly procurement policies, the budget of the case organization has apparently not been adversely affected since the inception of the reform. Respondents from the case study and the school system believe they continue to enjoy considerably generous financial and equipment allowances since inception. Perhaps this belief can explain the reason why the need for input measures, or information about cost efficiency in general, is not referenced by any of the respondents to this research.

This is quite an exceptional situation in a public sector world that typically has to deal with budgetary restrictions on continuous basis. The situation also allows us to explore, probably for the first time, the claimed importance and attributions of unremitting commitment of resources. If the experience of other public organizations that have experimented with performance measurements highlights the importance of committing resources then the case study organization should be in a position that allows it to harvest the upshot of a facilitating factor, abundant financial resources, as identified in literature and practice.

However, findings from this research cast some doubt on the power of the financial resources factor, at least as a single factor, over data utilization. **A prevalent lack of qualified human resources**, as reported by most respondents and supported by the archival review, seems to play a more important role in influencing measurement data utilization. Committing financial resources is apparently not enough to compensate for lacking human resources; the negative effects of human resources scarcity seem to override the potentially positive effects of abundant financial resources. As one respondent from the case study organization summarizes the situation: *"we are simply unable to find qualified people to fill our positions despite the very competitive salaries we offer"*.

Many respondents from the case study organization confess that they don't have the required skills to collect adequate data or conduct proper analysis of the collected data. Interestingly, some even acknowledge their confusion and lack of

understanding of the ideas behind establishing a performance measurement system in the first place. Likewise, almost all respondents from the school system admit that they don't have the knowledge and technical know-how to use the measurement data.

While the literature focuses on the importance of having qualified staff, with the right technical knowledge, who are devoted to collection and evaluation of performance measures inside an organization experimenting with measurements systems (Holzer and Halachmi, 1996; Wilkins, 1996), this research highlights the importance of having 'qualified' audiences and recipients of measurement information outside the organization. The issue around qualified human resources extend to the 'end-users' of the performance measurement information who are expected to utilize this information in order for the reform to work efficiently. In describing her experience with data utilization, one respondent from the parent group reports that: '*they want us to drive a Rolls Royce when we don't even know how to ride a bicycle; they need to put the people involved in this reform through a school that teaches them the ABCs of reading and understanding data in these reports they keep sending us*'.

Moreover, the human resources situation in the case study and the school system is further complicated by a sudden introduction of a policy that mandates a high percentage of nationals at the reform's workforce. At the beginning of the reform, the pool of potential staff, teachers, and administrators was reportedly open to regional and international candidates. The new policy is regarded by some respondents from the case study organization as premature since it has taken place at a point of the reform where required skills are not yet available at the national workforce level. Although respondents from the case study management cadre admit that the new policy is not strictly abided by in their hiring practices, they acknowledge its implications on the behavior of their local employees:

"Employers are fighting over qualified nationals who typically have job offers lined up in front of them. If a local employee is unhappy here, they can step out of our

building and find another job in a matter of hours"… "Local employees are taking
full advantage of these policies; they have no incentive to be trained…why would
they if their jobs are guaranteed"

(Respondent from the case organization)

The collected data for this research reveals an ongoing debate about national
capacity and the best methods to build this capacity. Training is recognized in the
literature as one of the ways to acquire the technical knowledge of how to conduct
and implement performance measurement which is critical to the success of these
systems (de Lancer Julnes and Holzer, 2001). There is scattered reference to the
importance of creating a centralized training program to address human resources
scarcity. However, not all respondents agree to such an approach and many think
that the nationalization policy must be reversed to allow access to a much needed
international expertise. In the absence of a clear direction to address current
limitations, the human resources factor, along with accompanying restrictive hiring
practices and unsatisfactory capacity building programs, seem to a be key
influencing factors that affect the utilization of measurement data.

This finding is also consistent with other research that warns against a worldwide
plethora of performance measurement systems (Power, 2000) which was not
necessarily accompanied by the kind of capacity building required for implementing
and using the data that these systems produce (Hatry *et al*., 2003).

5.4.2. Compatible organizational set-ups

While the findings of this research support the literature position regarding the
importance of having devoted qualified human resources with the right technical
know-how about collection and evaluation of performance measures, it also brings to
the discussion **an emerging theme related to organizational readiness from a
structural, policy, and procedures points of view**.

Interviews and archival review reveal that the case study **has not developed or introduced any policies or procedures to govern the functioning of the performance measurement system** and its relationship with other functions and units within the organization. There are no internal policies within the case study organization regarding the utilization of measurement data. Neither are there procedures in place to guide or influence processes related to data analysis, reporting, and utilization. One respondent from the case study comments that *"It's like putting a lot of effort to develop a system without thinking about how it will be run; what procedures to follow, who will be in charge of data analysis and reporting"*.

Moreover, and up until 2007, there was not a single position in the organizational structure that is dedicated to overseeing the performance measurement system. Although new positions were created in 2007 to allegedly monitor and improve the utilization of measurement data, these positions were still not functioning up to the writing of this thesis. The respondent from the policy maker group acknowledges the importance of staffing and operationalizing these new positions, *"otherwise no one has the time, or motive, to even look at the produced data."* Perhaps a more thoughtful and deliberate approach to the required organizational arrangements might address some of the data utilization challenges that face the reform. The development of policies and procedures that ensure implementation and utilization of the measurement system and its outcome data should probably be regarded as a prerequisite that can potentially allow for, and lead to, staff devotion to evaluation and utilization of performance measures. Similarly, relevant positions can be one of the means to guarantee the staff devotion to data evaluation and utilization referenced in the literature.

5.4.3. Supportive legislative set-up

Another emerging theme, one that is not addressed in the interview questions and probes, is related to the **lack of legislative mandates or external requirements** to

support the newly established performance measurement system and endorse the utilization of its outcome information. Three out of the eleven interviewees made reference to the importance of officially regulating the implementation of the performance measurement system by external parties in the government.

Although external requirements are considered in the literature as a rational construct that can potentially affect utilization (de Lancer Julnes and Holzer, 2001), the researcher originally assumed that the top down nature of the reform, in and by itself, will act in the capacity of an external requirement substituting the need for further external requirements. This finding highlights that this assumption is not accurate as respondents clearly point to the existence of a gap in this area.

"Many countries have issued laws that specify how measurement data must be used…You can't leave the usage and reporting on findings up to people's whims in the public sector; this is an important activity that needs to be regulated and monitored by the government"
(Respondents from the case organization)

As discussed in the literature review chapter, there has indeed been a noticeable increase in the introduction of mandates and frameworks around the world to monitor the implementation of performance measurement systems. Examples include the 1993 GPRA in the USA, the issuance of the 1999 VBTB in the Netherlands, and the 1992 establishment of the Office for Standards in Education (OFSTED) in the UK. Perhaps the experience of the case study indicates it is time to move in that direction.

"I now strongly believe that without the legal teeth it's unlikely that collected data will ever be used…neither will employees ever feel obliged to share, analyze, discuss…the results. The country should start thinking about introducing such steps"
(The respondent from the policy maker group)

It is worth noting that the literature, in assessing the decisive influence of internal and external requirements on performance data utilization, points to availability of resources as a key reason that can affect data utilization in a context where requirements exist (de Lancer Julnes, 2001). On the contrary, the availability of financial resources at the case study organization is not proving to be a key reason that can induce or affect data utilization in a context where requirements do not exist.

Summary

The experience of public organizations that have experimented with performance measurements highlights the importance of committing resources to the success of performance measurement systems implementation. Findings from this research reveal that the case study organization is not able to harvest the benefits of this facilitating rational factor due to overflowing negative effects of other rational factors.

In particular, **the research strongly supports the literature about the importance of qualified and technically knowledgeable staff who are devoted to collection and evaluation of performance measures inside an organization experimenting with measurements systems.** Clashing human resources policies at the national level are exacerbating the situation and adding to the challenge. Two emerging themes point to the influence of **unaccommodating organizational setups and lack of legislative requirements** on measurement data utilization. The absence of internal policies and procedures and external mandated requirements are reportedly influencing the poor utilization of measurement data. Figure 5.4 provides a summary of the different rational factors that are explored in this research and their claimed influences on utilization.

Financial Resources	→	Facilitating Factor	→	Influence not established
Human Resources	→	Inhibiting Factor	→	Reported Influence
Organization-al Setup	→	Inhibiting Factor	→	Reported Influence
External Requirements	→	Inhibiting Factor	→	Reported Influence

Figure 5.4: Influences of Rational Factors on Performance Data Utilization
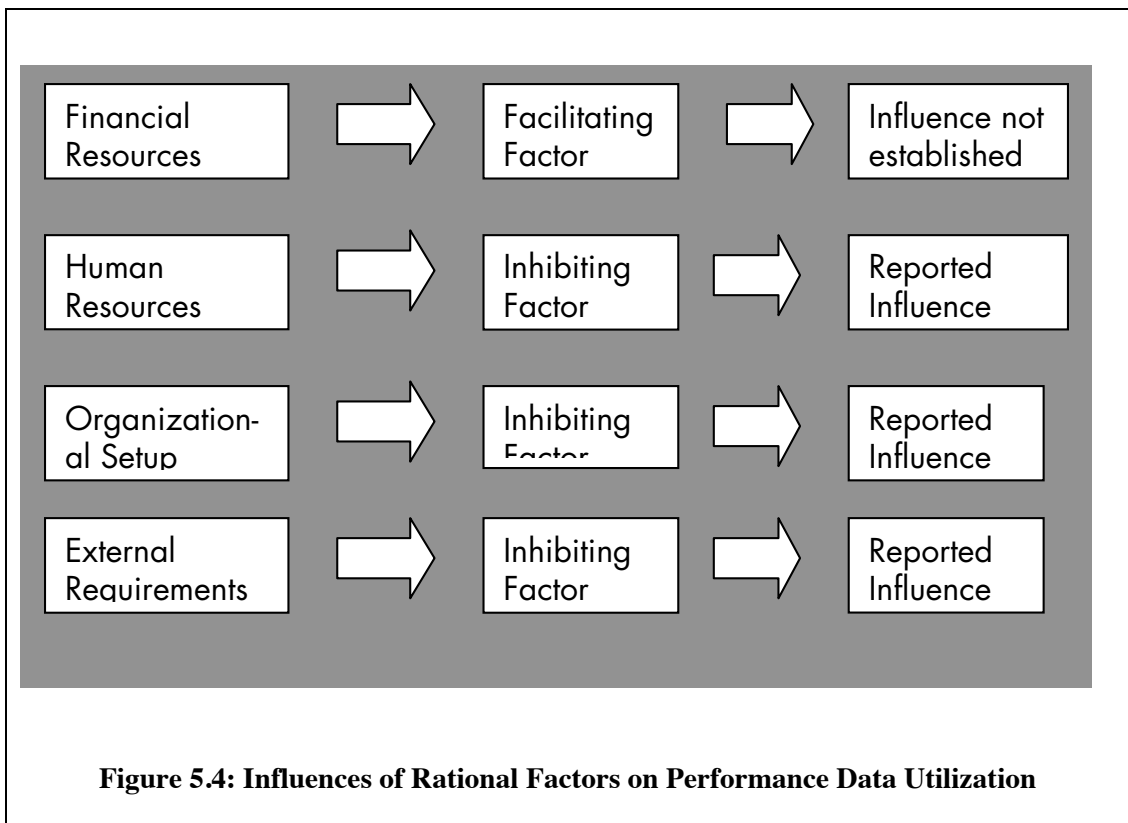
5.5 Sustained Leadership Commitment is Crucial to Move Forward a Progressive Reform Agenda

This research explores key factors that influence and affect utilization of measurement information from the perspective of key stakeholders. One of the main findings of this research indicates that the level of political will, support, and commitment, as perceived by key stakeholders, has a definite influence on the utilization of the produced measurement data and information. **Seven out of the eleven interviewees, and all school administrators, teachers and parents**, make direct, unsolicited, reference to the role that demonstrated leadership support plays in shaping their opinion and position regarding the reform in general and the measurement data in particular.

This notion about political support and commitment constitutes an emergent theme in this research. Despite the fact that it has long been recognized as a key factor in reform efforts (Nunberg, 1997, Campos and Esfahani, 2000), the original set of questions used in interviews and the focus group do not include any reference to the topic of political commitment. This is mainly due to the orientation of this research that distinguishes between the stages of adopting and initiating a public reform and that of utilizing the outcome information of its measurement system. Thus the focus of this research, and the questions posed in the interviews and focus group, is on the utilization of measurement information versus overarching political factors related to the introduction and adoption of reform initiatives. However, this research's data review suggests that **the effect of political commitment extends beyond general adoption themes of reform initiatives into actual implementation and employment of adopted policies; namely on- the-ground utilization of measurement information by different stakeholders.**

To better understand the issues related to the influence and effect of leadership's support in reform settings, and given that this theme emerged out of the collected data, the researcher reviewed a number of studies in the literature on the subject during and after the data analysis stage (Heaver and Israel, 1986; Nelson, 1990; Johnson and Branson, 1993; Williamson, 1994; Jayarajah and Branson, 1995; Nunberg, 1997; Killick, 1998; Campos and Esfahani, 2000; and McCourt, 2003). The researcher was looking for studies that examine the question of what makes governments implement an espoused policy. In particular, studies that test the relationship between political commitment and successful implementation of reforms. Among the list of factors typically associated with the overall success of reforms, visionary leadership (Williamson, 1994), expression of political will by top leadership (Johnson and Branson, 1993), executive authority (Nelson, 1990), and leadership (McCourt, 2003) surface as key features of successful reforms. The relationship between leadership's commitment to reforms and, at least, the successful initiation of reforms is seemingly well established.

However, and despite the evidence in the literature, it is important to bring attention to the fact that most reform cases that are examined in the literature share a number of similar themes and characteristics. Most of the studies on political commitment to reform, which generally fall into the areas of enterprise reform (Campos and Esfahani, 2000), civil service reform (Nunberg, 1997, McCourt, 2003), and macroeconomic management (Killick, 1998), apply to developing countries that depend on donor aid and foreign funding to initiate reform. Many of these studies are concerned with factors associated with reform success and warn against artificial or false political commitment at the inception of reform as means to guarantee funding.

As the discussion about the case study organization indicates, the reform in the case study is completely funded by the local government and enjoys strong financial support. It follows that it is safe to assume that the leadership's demonstrated political support and commitment during the inception of the reform is genuine or, at the very least, not instigated by a fund raising agenda as suggested in the literature. This point potentially has some bearing on the relevance of the current body of knowledge to this research. Perhaps some of the conclusions and recommendations suggested in a rather pessimistic commitment literature that endeavors to arrive at producing universal model of political commitment are not necessarily applicable or relevant to this discussion; at least not entirely.

Given that a large subset of the respondents who coin the leadership commitment theme in this research makes a clear distinction between what they think of as the leadership's 'initial' and 'subsequent' positions of the reform, the following subsections attempt to organize the findings of the data analysis into two eras: the early two years of reform inception, and, the years that followed.

5.5.1 Leadership Commitment during the Early Years of Inception

A review of **archival records** and publications (Appendix 1) depicts a picture about the early beginnings of the reform initiative that is characterized by the strong

engagement and support of the country's ambitious and forward-looking top leadership. In fact, the entire reform idea is conceived by this leadership supported by the primary consultant group as per archival records. To introduce and implement a top-down reform agenda, the leadership took a number of bold steps including major exemptions to the legal status and reporting mechanisms of a new organization, the case study, which is established to lead the reform.

During this early period of time, 2003-2005, the reform is reportedly protected from major backlashes by a leadership that publicly buttresses adopted policy changes and expresses its support to what is transpiring in the country. Data from interviews, focus group, and media records portrays a leadership that is visibly and actively participating in relevant events during this period of time.

"If it wasn't for the leadership's full support, political will, and clout, this reform project would have never seen the light of day. They 'the leadership' simply made it happen".
(Respondents from the primary consultant group)

"This reform was adopted and introduced by the country's leaders; they saw the need to improve and modernize education and they did everything possible to move in that direction. No other party can play that role here. The leaders provided political support and the required logistical and financial support to turn ideas into reality"
(The respondent from the policy maker group)

At the same time, collected data shows a highly responsive audience that is seemingly in tune with the attitude and beliefs of its leadership. In particular, collected data shows that **local** stakeholders such as local members of the management cadre, the media and the focus group, take their cues from their leadership's demonstrated behavior towards the reform. The beliefs and attitudes of these local stakeholders towards the reform in general, and the measurement system

in particular, are strongly influenced by what they believe to be their leadership's positive position towards the reform's components.

"Our people look up to their leaders for guidance and direction; if they (leadership) think this is a good system for us rest assure we will use it"…"They decided we need to make a shift to using measurement to improve educational decisions; they must be right and we should all trust their judgment"…" We follow our leaders' directions. They know what is best for us and they support us"
(Respondents from the school system group)

However, if established leadership commitment is indeed what it takes to induce measurement data utilization by stakeholders, then what explains a reported poor utilization of this data? If citizens do take their cues from their leadership, as the collected data shows, then why do almost all respondents to this research report their consistently weak, or non-existent, utilization of the measurement data, even during a period perceived to be characterized by notably strong leadership commitment?

There are clearly other factors at play. Certainly the first finding in this chapter offers an expansive discussion about the patterns for using measurement data in strategic and operational decision making, and the extant reasons behind it, by **interviewees** from the case study organization. These reasons vary from irrelevant mix and focus of measurement data, to tardiness and poor analysis of this data**. Some of these reasons might be relevant to a discussion about measurement data utilization, mainly by internal stakeholders, during the early period of reform inception.**

Also, a further examination of data collected from the **focus group** reveals yet another reason that can potentially be affecting their utilization process during that early period of time. The reason is chiefly related to a **general lack of experience and training in information interpretation and inferring meaning from reported measurement data** among this group of external stakeholders.

Moreover, the fact that the performance measurement system was still in its infancy during the first years of inception is likely to be another influencing factor. The earliest public announcement of measurement data took place more than a year after the initial implementation of the reform according to archival records and interviews. This simply means that **most of the measurement data was not publicly available for external stakeholders to use during that period; even if they were willing to start utilizing the data.**

"*Even though we administrated tests and started collecting data during the first academic year of the new reform schools, it took us until the middle of the next academic year before we were able to publish this data. Before that, it was mostly available for our use within the 'case study' organization*"
(Respondents from the management cadre at the case study organization)

While the findings of this research reveal a number of existing factors that can allegedly offer themselves as potential reasons behind poor utilization of measurement data during the early years of inception, **the case for the effect of leadership commitment, the topic of this finding, on data utilization is not quite made despite its reported significance and direct influence on the utilization decision** (Figure 5.5). Perhaps it is conceivable to argue that strong leadership commitment could have triggered a utilization process, mainly amongst external stakeholders, if measurement data was fully available to these stakeholders, and, also maybe, if these stakeholders were offered a good level of training on how to use this data. Under this scenario, poor availability of measurement data and lack of knowledge about data interpretation act as inhibiting factors that can override the effect of strong leadership commitment.

In addition, and based on observations about the behavior of stakeholders inside the case study organization, it is plausible to argue that strong leadership commitment, even when measurement data is available, is not enough to induce stakeholders to utilize the data. In this scenario, other inhibiting factors bear more weight than leadership commitment onto the decision to utilize measurement data.

In all cases, it remains difficult for this research to speak with any certainty to the presence or type of a potential relationship between positive political support/commitment and actual utilization of measurement data: a major upshot of the reform's new policies. It is not possible to make any inferences from the collected data regarding the two, seemingly at odds, phenomenon that characterize the early years of inception: strong political support of the reform and poor utilization of measurement information by stakeholders.

Figure 5.5: Unclear Relationships between Strong Leadership Commitment and Poor Utilization of Measurement Data
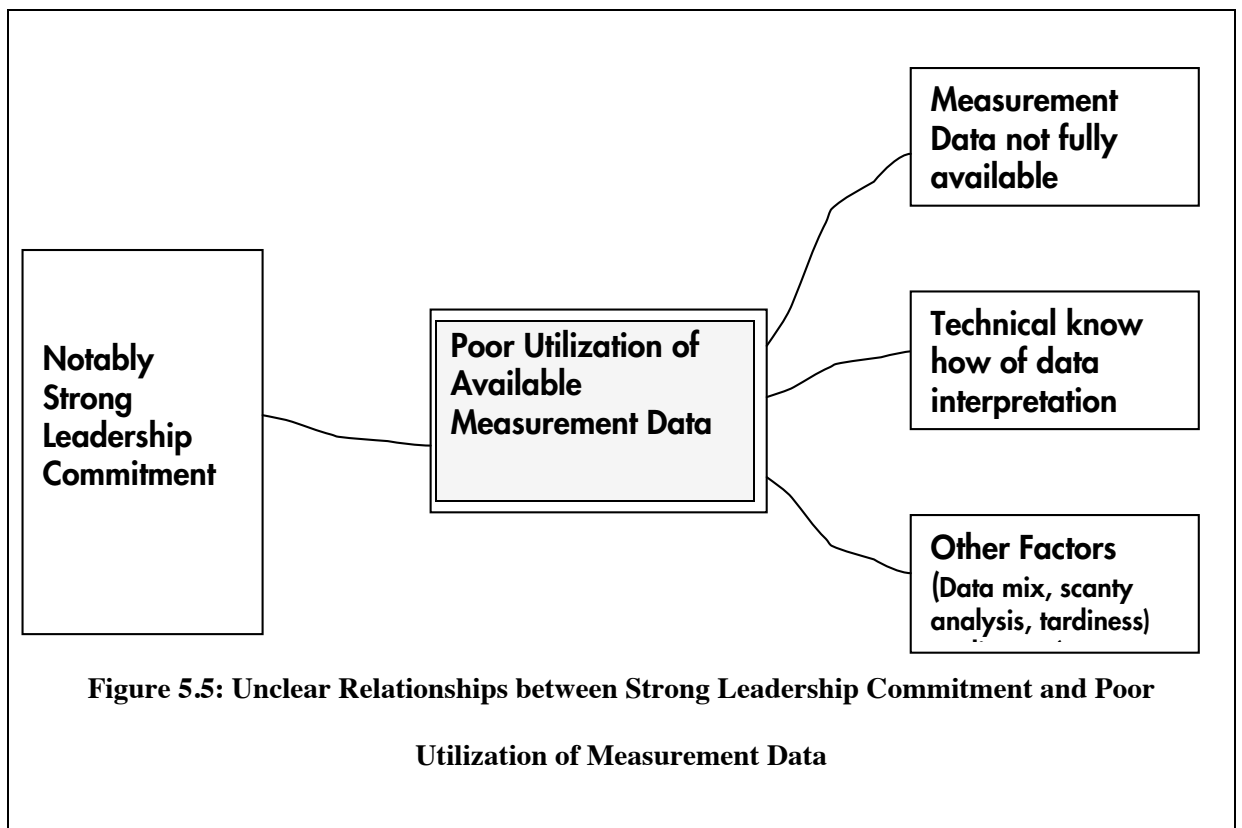
5.5.2 Leadership Commitment during the Following Years

There is noticeable agreement amongst most respondents to this research about a declining trend in leadership support and commitment with the passage of time. In fact, a number of these respondents frequently use descriptions such as 'change in leadership attitude' and 'leadership's waning support' as part of their attempts to explain some of the reform's episodes and behaviors. The initial '*zeal*' towards the new reform, as described by a member of the senior management cadre at the case

study organization, '*has now been replaced by a more realistic view of the system and a lukewarm attitude towards measurement data and its importance*'. Most respondents cite evidence regarding the decline of leadership commitment in their **public behavior** towards reform. A general view concerning a leadership that is increasingly reluctant to make public appearances in support of reform is indeed prevalent.

Since this is a top-down reform that is initiated by the leadership, it is not surprising that utmost importance is attached to the public behavior of this leadership towards the reform. Certainly the reform literature emphasizes the importance of committing leadership to a reform initiative through offering them a chance to **publicly** express their support (Heaver and Israel, 1986). Likewise, and in his model that depicts political commitment to reforms, McCourt (2003) identifies '**public**' as one of five key elements that are associated with, and can be an indication of, strong political commitment.

The claimed power and effect of 'public' leadership commitment is present in the collected data in two important ways. As illustrated in Figure 5.6, **the clout and effect of 'publicly visible' leadership's support are thought to profoundly affect the positions and attitudes of both the local media and the public**. Respondents believe that one of the repercussions of poor public leadership support has been manifested in the media's negative position towards measurement efforts.

"Once their (leadership's) support started to dwindle the entire measurement system started losing its role … those opposed or skeptical, like the media, took advantage of this unexplained change in leadership position and increased their attacks on the reform"
(Respondents from the case study organization)

"Even the media will be more supportive if they sense a strong political will to enforce the reform: this is how you make things happen around here"
(Respondents from the school system group)

In addition, data analysis shows yet another serious repercussion of the perceived weak leadership public commitment. This is the loss of faith and spread doubts regarding the value of the measurement effort. An audience whose attitude and belief is reported to be strongly influenced by signals they receive from their leadership is no longer motivated to use a system that, in their minds, has lost the trust and support of this leadership.

"How do schools expect me to take the 'school articles' seriously when all the signs clearly tell us that the country's higher-ups have abandoned this system" ."It's embarrassing for them 'leadership' to go out in public and admit that the system isn't working properly; that's why they are just being silent about it"
(Different respondents from the school system group)

"Our leaders…can now see the problems with the system especially after hearing people's complaints…I think they simply decided to withdraw from the scene… clearly they no longer are convinced with the benefits of measurement"
(Respondent from the teachers group)

Figure 5.6: Perceived Effects of Leadership's Public Position
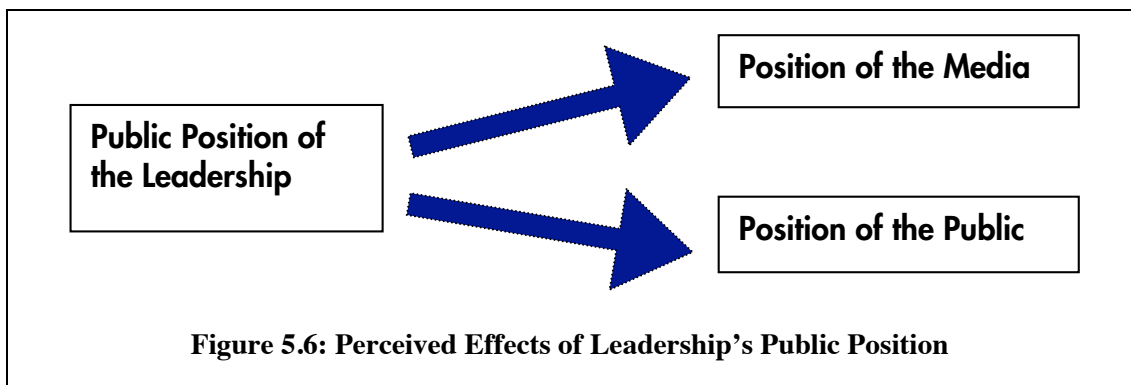
Conversely, a review of media records (Appendix 1) **does not fully support** the claim that **media attacks on the reform have increased over time**. In fact, the review shows a media trend that is rather sporadic in its features. The researcher finds it difficult to associate the random fluctuations in the media's positions over the years with reported observations regarding a declining trend of leadership support and commitment.

As a result, it is safe to state that claimed effects of leadership's public commitment on the position of the media are not substantiated in the archival review. At the same time, there is enough evidence in the collected data that indicates a relationship between leadership's public commitment and the beliefs and attitudes of the citizens. Figure 5.7 depicts a cycle that encapsulates a reported relationship between the different factors that are apparently contributing to poor utilization of measurement data.

An equally important finding is that the media review sparsely supports the widespread notion amongst respondents regarding leadership's declining **public** support of the reform. The review shows that the public visibility of leadership during reform-related events has only slightly decreased over the years.

However, it is not evident that these few, and somewhat isolated, observations constitute a statement from leadership about their commitment to the reform. There are certainly no reports or indications in the collected data that shows there are other signs of declining support. Respondents themselves don't offer any strong evidence to support their theory about declining leadership commitment beyond the poor public appearances account; as important as this factor might be. This then brings up a question about the accuracy of the public's reading of their leadership position from the reform in general and measurement in particular. It also remains a separate question whether the declining trend, if it actually exists, is a purposeful weaning practice from the leadership end or is indeed a detachment, or even abandonment, statement.

Figure 5.7: Established Relationship between Poor Leadership's Public Commitment and Poor Utilization of Measurement Data
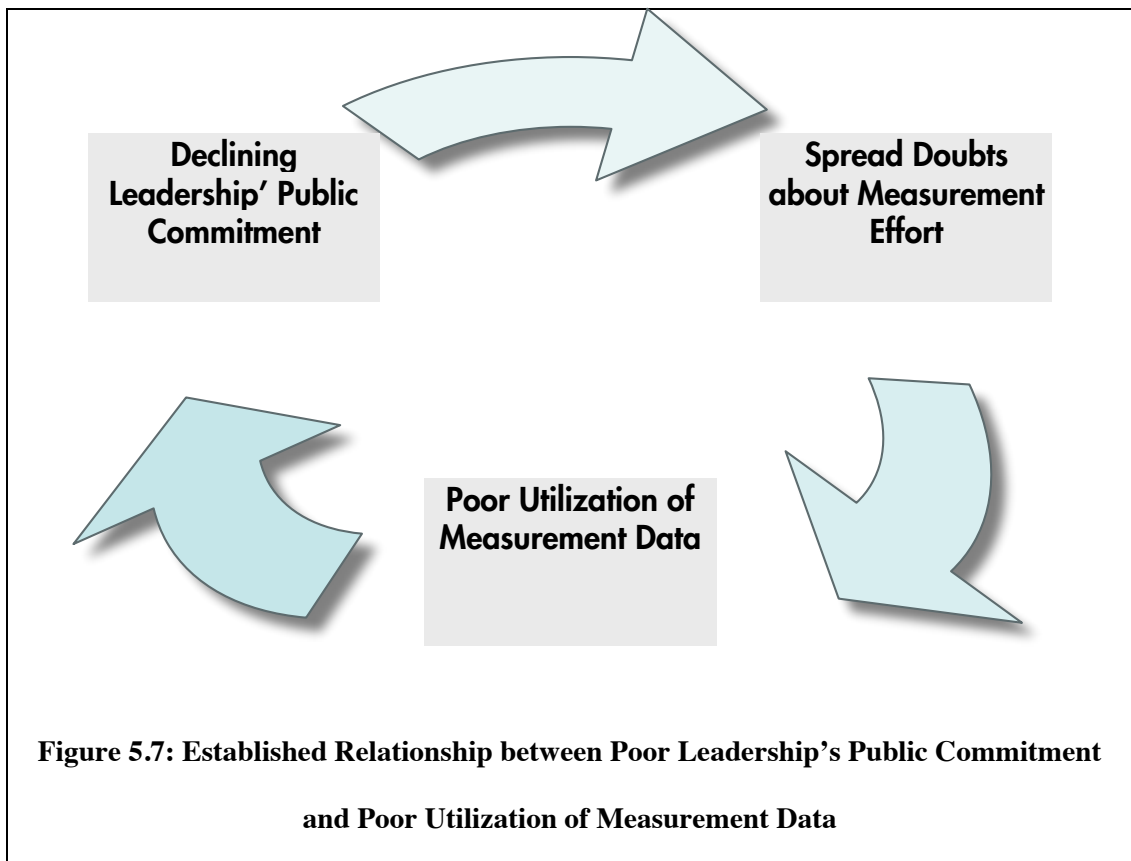
Summary

Political will and commitment to reform by a country is regarded to be '*almost universally recognized as one of the main factors explaining success … and is one of the most commonly quoted causes of unsatisfactory project completion*' (Heaver and Israel, 1986). Evidence on the importance of commitment to reform seems to be indeed compelling (McCourt, 2003). It is therefore quite surprising that the concept of commitment remains fuzzy, 'conceptually elusive' (Jayarajah and Branson, 1995), and used with different connotation in the literature. The use of the term 'commitment' ranges from references to local commitment, leadership commitment, and at times ownership. This research finding is concerned with leadership commitment.

A clear distinction is made regarding leadership's commitment during the early years of the reform and the subsequent years. Most respondents to this research attach

great importance to what they perceive as their leadership's position of the measurement system. Respondents also report a strong effect of what they think of as the leadership's position on their own beliefs and attitudes towards measurement data and, ultimately, on their decision to use this data.

Interestingly, judgments on leadership commitments are mainly passed based on public and visible appearances of this leadership during events related to the reform. Whereas the reform literature acknowledges publicly expressed leadership support as one of several other factors associated with commitment, this research **shows the 'public' factor of commitment as the key and most important indication of strong leadership commitment** in respondents' minds.

In addition to the top-down nature of the reform, another explanation is perhaps related to a noticeable absence of official public announcements from the leadership regarding the subject. The archival review shows no recorded evidence of leadership's commitment, or even reference, to the subject of measurement. In the absence of such guidance, citizens are apparently making their own inferences based on what evidence they can see and monitor. Such evidence is hence limited to the frequency of leadership's public appearances in reform-related events. Yet another explanation maybe related to the legal status of the case study organization. In the absence of legislations to enforce utilization of measurement data, political commitment seems to be the key method to encourage, or even enforce, a practice that would otherwise be viewed as discretionary at best if not an added burden.

5.6   Scrupulously Developed Internal and External Communication Strategies Are Required to Facilitate the Introduction and Acceptance of a New Measurement Culture

Public reform practitioners have seen enough evidence over the years to show them that communication is critical to generating broad support for reforms (Cabanero-

Verzosa and Garcia, 2009). The case for the importance of communication in a public reform setting is probably hard to argue against. Williamson (1994) lists "reformers making good use of the media" as one of 16 hypotheses that determine the success of public reform initiative. Williamson also lists 'social consensus as a powerful factor impelling reform' as another one of the 16 hypotheses. Arriving at social consensus requires, among other factors, the utilization of sound communication strategies.

Despite the overall agreement in the literature about the general role of communication in a reform setting, challenge seems to exist in answering the question of 'how' to communicate the reform message; what strategies and approaches should be used by those in charge of a reform communication program. A review of interview data clearly confirms the previous statements. **Almost all respondents** in this research make some kind of reference to the issue of the media and the important role it plays in this context.

There is also enough evidence in the collected data about an overall disapproval of the reform's current communication strategies and their associated messages. **Many respondents** from both the **case study organization** and the **focus group** believe that a more carefully designed communication strategy could have steered away a lot of the negative criticism that this reform has been subjected to since its inception.

In addition, and similar to the second finding of this research which highlights the distinctly different data needs of the different stakeholders' groups, this finding also highlights a rather **wide range of needs and expectations** from the reform's communication effort **even within the same group of stakeholders** in some instances.

A preliminary data analysis shows that stakeholders expect communication to (1) create positive awareness and educate the public, (2) shape and even create the new identity of the case study organization, (3) institutionalize new features of the measurement system and cement inter-organizational bonds through proper internal

coordination, and (4) showcase the new system at the regional and international arena. However, further, and more scrutinized, review of the collected data shows that certain groups of stakeholders harbor more harmoniums expectations while members of other groups of stakeholders have almost opposing views regarding the reform's 'ideal' communication approach. In order to anchor some of the emerging themes in substantive material, further review of relevant literature has been conducted by the researcher during and after the data analysis stage.

It is noteworthy that much of the relevant research in this area is done by development agencies such as the World Bank. The work done by Mefalopulos (2008) on communication models is selected as a reference point to discuss this research's finding because of its comprehensiveness and relevance to a reform setting. The model, summarized in Table 5.1, describes four key types of communication strategies that are commonly used by organizations in charge of reform initiatives. The suggested communication strategies are highly complementary but different in scope and function; they are not mutually exclusive and can overlap.

Type	Description
Internal communication	Ensure timely and effective flow of relevant information between management and staff regarding the policies and practices of the organization. Promote cohesive organizational structure needed to conduct the organization's business effectively
Corporate communication	Promote the organization's mission, explain its programs and activities, and create a brand and position for the organization in the minds of its constituents and clients. Mostly for external audiences
Advocacy communication	Influence audiences and support the intended change, articulate the organization's stance on reform issues, and facilitate dialogue on critical reform challenges.
Development communication	Support sustainable change, enhance the effectiveness of reform by engaging stakeholders throughout the design, implementation, and monitoring of these interventions. Induces behavior and social change

	by establishing conducive environment

Table 5.1: Four Types of Communication Used by Reform Organizations

(Source: Mefalopulos, 2008)

In order to facilitate the discussion about this finding, collected data from interviews and the focus group is divided into two main categories: (1) feedback regarding internal communication efforts within the case study organization and, (2) feedback regarding communication efforts that target external stakeholders outside the case study organization. Internal communication is defined here as all kinds of messages and information that flow within the case study organization to introduce, explain, communicate, and highlight features and characteristics of the performance measurement system. External communication covers all similar efforts that target stakeholders outside the case study organization; inside and outside the country.

5.6.1   Internal communication/coordination

Guided by the literature review, the questions used during the interviews and focus group probed deeply into the issue of internal communication and asked for specific information about the exact communication venues currently used to share information produced by the measurement system. Interestingly, the majority of respondents within the case study confirm that almost none of the internal communication venues suggested in the literature are systematically used, or even used at all, to disseminate information in their organization (e.g. email, formal and informal meetings, and briefings).

The collected data reveal a predominant agreement amongst **public officials in the case organization** about the **absence** of effective and reliable **intra-organizational communication** to explain and educate staff about the measurement system itself; its components, processes, mechanisms, and outcomes. At the same time, many of these respondents recognize the damaging consequences of missing the opportunity

to inform internal staff about the characteristics of the new system and turn them into advocates of measurement in the society. In a relatively small-sized organization, even the policy-maker and some directors at the case study are not fully aware of the components of the measurement system and what these entail and produce in terms of data and information. It is quite interesting to note that a major measurement activity that has been introduced by the Assessment Entity a year prior to the interviews is virtually unknown to a number of public officials at the case study.

"Before we worry about changing the attitude of the society let us first worry about changing the attitude of our own staff, most of whom don't even understand the basics of measurement and the purposes it can potentially serve"…."Each one of us should be a proponent of the new system. However, it takes some serious communication effort to turn our inside people into believers"
(Different Respondents from the case study)

In addition to **educating** internal staff about the components of the system and **creating awareness** about its overall characteristics and advantages, interview data reveal yet another aspect related to internal communication and coordination. Many respondents from the case study organization are concerned about not sharing the results of the measurement system internally within the organization prior to their publication. Most organizational units, as well as the policy maker, are first notified of the system's results through newspaper articles and other public media announcements.

As a consequence of this absence of internal coordination, the organization is losing the opportunity to **internally digest the information** and prepare itself to deal with external reactions to published data. There is no time to **articulate a harmonious, and unified, organization-wide position** on matters related to measurement results. Bearing in mind the progressive and at times rather controversial nature of the introduced changes, lack of internal communication is clearly not helping the overall position of an organization in charge of introducing and defending the reform.

"It's rather strange; collected and publicized data is transmitted out of our organization to different external recipients without going through an internal cycle (either before or after transmission). It's like we are not concerned with the findings, it's like this is not our system"…." This entire measurement effort feels like an artificial 'limb' that is not connected to our body; we are detached from it, we don't understand it, we have conflicting opinions about it"
(Different Respondents from the case study)

If the case study organization is to advance new measurement culture and practices concerned officials need to 'make them comprehensible and enlist the support and cooperation of those who must work together to produce the intended result' (Moore, 1995). It follows that the organization's mission, its programs and activities, and the new principles of measurement must be well understood and accepted by the internal staff of the case study organization. Indeed a strong comprehension of the reform can be regarded as a pre-requisite to ensuring the buy-in and support of internal staff to the introduced change; maybe even a pre-requisite to trying to influence the attitudes and opinions of external stakeholders as some respondents argue.

Efforts that promote an organization's mission and explain its programs and activities are categorized as **corporate communication** by Mefalopulos (2008) who views it as mostly targeted at external audiences. However, **this research shows that corporate communication is quite fitting for internal audiences** in a reform setting; especially during the early stages of a reform. Moreover, if internal communication is indeed, and as defined by Mefalopulos (2008), a way to promote the cohesiveness of an organization allowing it to conduct, in this case, its reform business effectively, then the case study organization is compromising its ability to effectively carry out the responsibilities of the reform. The apparently blemished internal flow of information, characterized by weak internal coordination practices, seems to have resulted in poor comprehension, acceptance, and ownership of the new measurement system inside the case study organization.

Communication Approach	Expected Outcomes	Mefalopulos's Equivalent Type
Educate internal staff	Staff more supportive of reform. Turn staff into advocates	Corporate Communication (Not fully applicable)
Ensure effective flow of relevant information about the system among staff	More cohesive organizational structure needed to conduct the reform. Unified organization-wide position	Internal Communication

Table 5.2 Suggested Approaches for Efficient Internal Communication Strategies

(Source: Data from Interviews with the Case Study Officials)

5.6.2 External communication

The respondent from the **policy maker group** and **one** of the respondents from the **senior management group** at the case study organization share a conviction that poor internal communication practices contributed, at least partially, to the development of poor external communication strategies and plans. An overall misinformed and uncoordinated organization is, allegedly, ill-prepared and incapable of developing efficient external communication messages. Although this view is expressed only by the two respondents identified above, it brings to light an interesting point about the intertwined, and possibly chronological, relationship between internal and external communication strategies.

Perhaps a reform organization that lacks the cohesive internal position required to conduct its mission, an organization that is yet to create a distinct brand in the mind

of its internal staff, is not in a good position to convey its programs to its external constituents.

Even for respondents from the case study organization who are not able to explicitly articulate a relationship, the implications of poor internal coordination come to the forefront of their discussion about external communication efforts. The following excerpt captures the thoughts of the head of the communication unit at the case study organization on this matter:

"It's an endless cycle; we are constantly blamed for not issuing strong communication messages in time to address negative public reaction to measurement data. How can we produce an appropriate (and unified) message when we get the results at the same time these are shared with the public…no wonder we are always caught up in reactive crisis management modes"
(Respondent from the Senior Management Cadre)

5.6.2.1 Respondents from the case study organization in relation to external communication

Apart from references to the relationship between internal and external communication, data analysis shows that most respondents in this research have their own views regarding efficient external communication strategies. At the same time, these views and expectations are quite different and vary even among the same stakeholder group.

A good example is the variation of input received from public officials at the case study organization. Some of these officials hold strong opinions in favor of focusing external communication messages on featuring and **showcasing** the modern and advanced measurement system at the local, regional, and international levels. The goal of this public relations-like approach is not to educate the public; rather it is to draw the public's attention, inside and outside the country, to this new system and

showcase it as a national pride. As shown in Table 5.3, **this focus is not captured in any of the four communication strategies** described by Mefalopulos (2008).

"Because we failed to show the world what a great reform this is, most of our efforts fell into obscurity. This is the first measurement system of its kind in the region, but a very limited circle of people are aware of this"… "During international conferences and events, I meet experts who lecture about advanced measurement systems and their advantages to governments and the public sector. We have such a system but no one is aware of that. No one can see us because we did such a bad job making others aware of what our organization has to offer. It is really de-motivating".
(Different respondents from the case study organization)

At the same time, there is another group of respondents within the case study organization that feels strongly against adopting a communication approach that promotes showcasing the system and focuses on 'cosmetics' as one respondent dubs efforts to show off the system inside and outside the country. This group stresses the importance of focusing external communication messages on **educating the local public** and **making the system understandable** by the people who should be using it at the end of the day. They believe that the success of the reform hinges upon the public practice of certain behavioral patterns. Consequently, they believe that all communication efforts should focus on ensuring that the public understands the system in order to utilize it.

"A good portion of the criticism we receive from the local media could be eliminated if we only succeed in making people in society understand the system and how it serves them".. "Instead of wasting time and money bragging about what advanced systems we have, all of our messages should focus on educating the public… this is what should come first; propaganda and meaningless cosmetics can wait few years down the road". "This entire reform is built on transparency and accountability. It will crumble if we fail to get the public to interact with us and take decisions based on the information we provide them"

(Different respondents from the case study organization)

As shown in Table 5.3, this focus on educating the public and explaining one of the reform's key components, measurement, is **consistent with corporate communication** as defined by Mefalopulos (2008). Media outputs and products are used to promote the mission and values of the reform institution; selected audiences are informed about relevant activities.

However, this discussion has another key dimension in addition to the purely educational one. It raises attention to a perceived risk of obliterating the entire reform if the 'wrong' communication strategy is adopted. A 'wrong' communication strategy is apparently one that does not pay enough attention to educating external stakeholders about the reform. Indeed, the literature shows that there is a body of evidence that has emerged since the 1980s from reports of major international aid organizations showing that reform efforts often flounder because local people are left out and not involved in the programs (Rahnema, 1993). It is interesting to note that respondents to this research only bring up fears about a possible collapse of the reform during a discussion about communication. For example, the elaborate discussion about stakeholders, under point 5.3 above, carries no reference to perceived risks associated with selecting 'wrong' performance measures that are not in line with the needs of stakeholders. The fact that respondents associate risks regarding the collapse of the reform with external communication strategies could be a testimony to the perceived power of communication and the role it can play in changing the attitudes of the public.

Communication Approach	Expected Outcomes	Mefalopulos's Equivalent Type
Showcase the measurement system inside and outside the country. Public relations	System becomes a national pride. Case study gains national and international recognition	Not Applicable

approach		
Educate the public	Better support for the reform	Corporate Communication

Table 5.3: Suggested Approaches for Efficient External Communication Strategies

(Source: Data from Interviews with the Case Study Officials)

5.6.2.2 Respondents from the school system in relation to external communication

On the other hand, focus group data reveals that respondents from the school system hold a more harmonious set of views regarding the reform's external communication efforts. Moreover, data indicates a suggested **phase-specific approach** where different strategies are utilized at different stages of the reform. These respondents claim that their daily interactions with parents and other constituents of the school system give them the advantage of identifying the real expectations and needs of the society on reform issues. They are noticeably less concerned with celebrating successes and reaching regional and international audience but **more focused on carefully addressing the needs** of the society inside the country. They believe in the importance of acquired knowledge in **shaping new beliefs and changing the attitude of the society** toward the new system in general and the new schools in particular.

Data review also reveals that most school administrators and teachers from the focus group are not impressed with what they think of as 'sophisticated and fancy' messages that are currently used in the reform. They believe that such messages don't resonate well with ordinary citizens.

"The editorials and ads put by the organization speak a language that is overly intellectual and hard to comprehend by many people in the public"…."Not many

parents are experts in measurement and standards, must find a way to win the hearts and minds of ordinary and simple people"
(Different respondents from the school system)

The observations of the schools' administrators and teachers are consistent with statements made by the parents in the focus group who also express a good deal of confusion about the messages concerning the measurement system:

"We are overwhelmed by all these reports that they send us and don't really understand what they mean"... "I can't claim that as a parent I can now make an informed decision about what school to select for my children. I don't understand what they send us and can't figure out which school is better than the other"... "The meetings with officials at the 'case study organization', and even the ads they place in the papers, are neither informative nor useful, I always end up being more confused every time I attend one of those public forum meetings"
(Different respondents from the parents group)

Apparently, creative and sophisticated messages, as artistically appealing as some respondents describe them, are not helping stakeholders comprehend the new culture of measurement. What seems to be missing is a clear understanding and appreciation of the practices, knowledge level, and beliefs of the different groups of stakeholders who are affected by the reform. This understanding can then guide the formation of the communication messages and better tailor them to the stakeholders' practices and knowledge levels. Current communication strategies at the case study organization are apparently not anchored in this understanding.

Another important finding of the data analysis is a suggested **sequential communication approach** for the reform to take root as shown in Table 5.4. In the minds of respondents from the school system, a reform's efficient external communication strategy should as a **first step, educate the society** on the new components of the system and how to use the information it produces. The society must be taught exactly how to use published data, how to interpret meaning from all

the numbers, and how to use this acquired knowledge to make better decisions about public schools. One teacher suggest the use of a series of TV episodes that provide practical and real life examples of collected data and the different ways to use this data. The idea is not to disseminate general themes about the reform, rather to provide detailed instructions on its programs and systems utilizing accessible and popular venues:

"Why not hire celebrities and use drama or even comedy to explain to the average person in the society what we are trying to accomplish in this reform…this is how you reach your masses; step by step easy instructions through popular media and using well-liked public figures that people resonate with"
(Respondent from the school system)

The suggested educational approach for the first stage of the reform is **consistent** with the definition of **corporate communication** as identified by Mefalopulos (2008): media outputs and products are utilized to promote the values and activities of the organization. The prevalent assumption among this group of respondents is that a better understanding of the system and its features, along with the newly acquired knowledge, will with time influence the attitudes and perceptions of the public about the reform. "*Obscurity leads to fear and suspicion*" as one school administrator points out, "*the public will only accept the system if they better understand it*".

The **next step**, according to respondents from the school group, should be an external communication strategy that focuses on **modeling new behavior** and **encouraging the society to practice this new behavior** that is based on their newly acquired ability to understand and utilize available information. Communication methods and media are to be used mainly to influence specific audiences and support the intended change in behavior.

In that sense, the suggested approach during this suggested second phase falls within what Mefalopulos (2008) calls **advocacy communication** strategy: the use of media to influence and support change at the public level.

"They (officials at the case study) should literally show people examples of how people act and behave once they get this information. Show them what they are supposed to do with their children and their teachers'…"I think actors should play out the role of well-informed and responsible citizens and show us the ideal kind of interactions that the reform aims to achieve"…"It will be a matter of time before the new reactions to information, taken initially by few pioneers, spread in the community and become part of the culture and normal practices"
(Different respondents from the school system)

The **last step** is suggested by a limited number of respondents from the school system who are concerned about the **sustainability** of the introduced changes. These respondents warn against the society falling back into old patterns of behavior in the absence of relevant communication strategies during a post-education, post-change stage. For these respondents, the role of communication does not end with the spread of new education and new beliefs and behaviors.

"It is very easy for our society to fall back into old practices; we have seen that happen time and again. If they want this reform to become more than a fad, they need to find a way to guard the changes and make sure they become part of our culture"
(Respondent from the school administrator group)

Although this group of respondents raises attention to the issue of sustainability, they are not able to articulate how a communication strategy can be developed to address this issue. Unlike the two other types of communication strategies that this group brings attention to, corporate and advocacy, respondents offer no examples of what methods this last communication approach can employ.

The closest equivalent to the suggested approach in Mefalopulos's model is the **development communication**. However, development communication requires engaging stakeholders throughout the design, implementation, and monitoring of communication efforts and intervention in order to support sustainable change. There are no suggestions in the collected data regarding continuous engagement of stakeholders even though there is a lot of reference to the importance of understanding the stakeholders' views and behavior.

Communication Approach	Expected Outcomes	Mefalopulos's Equivalent Type
Educate stakeholders, help then acquire new information/knowledge	Formation of new beliefs and attitudes towards the system	Corporate Communication
Encourage stakeholders to convert new beliefs to new behavior; promote new behavior patterns	Gradual practice of new behavior will spread and evolve into new social norms	Advocacy Communication
Maintain buy in and new behavior; support and reward new behavior	Sustainable change	Development Communication (Not Fully Applicable)

Table 5.4: Suggested Sequential Approaches for Efficient External Communication Strategies

(Source: Data from the Focus Group)

Summary

Reform initiatives require considerable effort and time to take root. There is overwhelming evidence in the collected data about a perceived critical role for communication to bolster efforts towards increasing the effectiveness of overall

reform efforts. Selecting sound and effective communication strategies seems to be regarded as a key policy and management decision for the success and sustainability of this reform.

It follows that thinking through communication alternatives and opting for the most fitting communication strategy should ideally be part of conceiving a reform initiative rather than an undertaking that follows the implementation of that reform. In making the case for the importance of early development of a communication strategy, Cabanero-Verzosa and Garcia (2009) argue that "when it is instead an afterthought used mainly as a means to disseminate information, managers jeopardize the potential long-term success of the planned reform". There is no evidence in this research's primary or secondary data indicating the existence of early, policy level, planning for a communication strategy to address the idiosyncrasies of this reform.

If successful reforms are indeed those that are understood by the general public and have the strong support of key stakeholders (Cabanero-Verzosa and Garcia, 2009), then this reform faces a fundamental challenge. A reported lack of comprehension and appreciation of the new measurement system, among internal and external stakeholders alike, seems to be posing a real obstacle in the path of securing support and buy in.

By the same token, the use of innovative and artfully packaged media messages has apparently not been sufficient to bring about the required changes in attitudes and beliefs. The current use of media mainly to raise awareness and inform the public about certain aspects of the reform, and transmit information about measurement results, is seemingly not producing the level of understanding required to appreciate and utilize the new measurement system.

Perhaps different communication approaches must be considered to achieve the required level of understanding and comprehension of the reform among stakeholders. Once again, the issue of understanding stakeholders' current levels of knowledge about measurement, practices, needs and perspectives comes to the

forefront of the discussion, this time for the purpose of developing better communication strategies and messages.

The data analysis strongly reveals that respondents promote employing a more stakeholder-centered approach to this reform's communication efforts. Participation of stakeholders and two way communications constitute an emerging paradigm in reform communication (Mefalopulos, 2008). The idea of engaging stakeholders from the outset of the reform, and working with them to assess the situation and devise effective strategies that lead to more sustainable reform initiatives, seems to be gaining more acceptances. A thoughtfully designed communication strategy must connect with key audiences in a way that increases their level of engagement and their commitment to sustained reform (Cabanero-Verzosa and Garcia, 2009).

5.7   Summary

This Chapter aims to present and discuss the findings from the analysis of secondary and primary data collected for this research. A reportedly poor, nonexistent, or nonsystematic utilization of performance measurement data and information can be regarded as a dominant behavioral characteristic among internal and external stakeholders who responded to this research. Consequently, the researcher decided to report mainly on key factors suggested by stakeholders as reasons that influence and affect their decision to use measurement data. These factors are grouped into major categories and presented in five key findings.

Producing measurement data that can be used to guide decisionmaking is one of the most recognized values of performance measurement systems. Perhaps in the context of a challenging, sector-wide, sweeping reform, more weight and importance are attached to appropriate and well guided decisionmaking given everything at stake. Due to the difference in proprietorship of the decision making processes at the case study, a distinction is made in the first finding between decision-making at the policy and strategic levels and that at the operation level.

Public officials at the policy and strategic levels perceive current performance measures as both inadequate and irrelevant to provide the governing body with the kind of information they need to follow up on the progress of the reform. Consequently, their decisions are reportedly rarely rooted in the performance measurement data. In particular, the technical concerns of the policy makers can be summarized as (1) design- focused, the mix and focus of the performance measures, and (2) human capacity-focused, the quality of the analysis and comparison with past data. Beyond purely technical and substantive concerns, two cultural reasons are also playing a role in hampering measurement data utilization by policy makers. These include poor cultural appreciation of data based decision making, and cultural tendencies to take decisions that are geared towards pleasing the public, regardless of whether the data supports the policy agenda of the reform.

At the operational level, public officials entertain similar doubts regarding the adequacy and capability of the measurement data to provide them with the kind of data they believe they need to carry out their mandates. However, the concerns of these public officials are strictly technical; political and cultural factors have no bearing on the poor utilization of performance measurement data in decisionmaking at the operational level. Technical reasons are also (1) design- focused, mix/focus of performance measures, frequency and tardiness of data release, format of files, and (2) human capacity-focused, the quality of the analysis and comparison with past data.

If performance measurement is intended as a means to make more informed decisions, then the first finding of this research offers a number of technical and cultural factors that are impeding this potentially valuable use of measurement data in the case study organization.

At a more overarching level, the research offers two intertwined factors as reasons that affect the overall perception of stakeholders of, and as a result their decision to use, performance measurement data. These factors are related to: (1) the role of foreign consultants in designing the performance measurement system, and (2)

unmet stakeholders' measurement data needs. While this research can not make a claim regarding a good understanding of the relationship between the two factors, it is clear that most respondents perceive some kind of an association between the two. On the wide spectrum of accounts offered to rationalize this connection, stakeholders' opinions vary from consultants' sheer negligence and lack of experience to hidden and purposefully malicious agendas of western consultants. Both factors date back to the performance measurement system design stage. The perceived failure of the consultants to address the complex issue of multiple stakeholders during the system's design and development stages, through the adoption of a unitary and over-simplified approach that focuses mainly on schools' outcomes, is playing a key role in the ultimate poor utilization of measurement data. A design that purposefully focuses on schools' outcome design can probably serve matters of parental choice and improvement of quality of schooling in the long run. However, it resulted in earnest limitations in areas that can be considered critical to the success of such a major reform. Moreover, problems of disagreements and incongruities between different stakeholders have only been pushed to the current policy implementation stage.

The second finding offers lessons about the importance of addressing stakeholders' needs early on during the design stage of the performance measurement system. The employment of a public sector performance measurement framework might have been helped address the voices of multiple stakeholders. Nonetheless, the notion of having to address several, and sometimes conflicting, stakeholders' requirements during the design of a performance measurement system has not yet been matched with research that clearly puts forward ways and means to systematically identify and analyze stakeholders and manages their relationships (Bryson, 2004; McAdam and Henderson, 2004). There is clearly a need for further research in this field. In exploring the impact of rational factors, borrowed from the field of evaluation, on performance measurement utilization, the third finding does not validate all the traditional effects noted elsewhere around the world. While the experience of most public organizations points towards the importance of committing resources to the success of performance measurement systems implementation, the case study

organization is not able to harvest the benefits of its abundant financial resources. This is partly due to overflowing negative effects of other rational factors. In particular, lack of technically knowledgeable staff who are devoted to collection and evaluation of performance measures inside the case study organization is playing an overriding negative effect on measurement data utilization. The negative effect of the human resource factor is further impaired by clashing human resources policies at the national level.

Two other rational factors surface in this research as factors that influence the utilization of measurement data. These are unsupportive organizational setups and lack of legislative requirements on data utilization. In the absence of internal policies and procedures, and external mandated requirements, data utilization is left to the discretionary impulse of an audience that is yet to overcome, or deal with, the many other obstacles to utilization cited in this research.

The influence of leadership support and commitment is found to extend beyond matters of reform initiation and implementation typically addressed in the literature. The fourth finding shows that sustained, and visible, leadership support has a profound effect on the beliefs and attitudes of stakeholders towards the use of performance measurement data. Local stakeholders in particular attach great importance to what they perceive as their leadership's position of the measurement system. They purposefully align their position, and ultimately their decision to use the measurement data, with that of the leadership's position, as they perceive it. Whereas reform literature acknowledges publicly expressed leadership support as one of several other factors associated with commitment, this research underscores the 'public' factor of commitment as the key and most important indication of strong leadership commitment in respondents' minds

Perhaps the top-down nature of the reform, accompanied by the absence of official public announcements from the leadership's side regarding the measurement subject, accentuates the weight that stakeholders' assign to the public appearances and endorsements of leadership. It is quite surprising that, in light of all the compelling

evidence for the importance of leadership commitment to reform, there is no documented evidence that portrays, or even references, the leadership's position on the subject of measurement which is quite critical to the success of the reform.

It is quite difficult to argue against the general importance of using sound and effective communication strategies to support any reform. However, the fifth finding shows that the current use of innovative and artfully packaged media messages to disseminate information is not sufficient to bring about the required changes in attitudes and practices regarding measurement data utilization. It highlights the importance of regarding communication strategies as a key policy and management decision for the success and sustainability of a reform. Thinking through different communication alternatives is suggested to be part of conceiving a reform initiative rather than an undertaking that follows the implementation of that reform.

If successful reforms are indeed those that are understood by the general public and have the strong support of key stakeholders (Cabanero-Verzosa and Garcia, 2009), then this reform faces a fundamental challenge. A widespread poor understanding of measurement culture among respondents puts the need to understand stakeholders' practices, needs, and perspectives at the forefront of the discussion once again; this time for the purpose of developing better communication strategies and messages. Employing a more stakeholder-centered approach in the reform's communication efforts is one of the key recommendations under this finding. This is in line with emerging paradigms in reform communication (Mefalopulos, 2008).

In the next Chapter the conclusions of this research are drawn and some recommendations are suggested based on these conclusions.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1   Introduction

The previous chapter provides a discussion of the key findings of the research. To demonstrate that the initial objectives of this research have been achieved, this chapter revisits the research questions and briefly report on how the findings from the study answer these questions in an explicit and unambiguous manner. The chapter then presents a summary of the key conclusions and suggests recommendations based on these conclusions.

To conclude this research, the chapter lists the contributions of the study, both in terms of its theoretical and practical implications. The quality and limitations of this research are presented followed by suggested areas for future research.

6.2   Answers to the Research Questions

The previous chapter provides a full and detailed account of the research findings. However because these are presented in an order that does not mirror that of the research questions, and to ensure that all the research questions have been covered and well addressed in the course of the research, this section provides a brief statement with the key findings for each of the research questions.

Research question number (1):

What happened to the data/ information that was collected and produced by the performance measurement system? How was it disseminated to the different stakeholders inside and outside the organization? What was done to share knowledge (emails, informal meetings, formal meetings, briefing, reports, and structured seminars)? Were there any changes that took place in the organization in order to encourage (or discourage) and/or institutionalize a process of utilizing the produced data/ information (training, new processes and revised Manuals of Procedures)?

The Assessment Entity in the case study organization is in charge of the performance measurement system. The entity collects data mainly through the administration of an annual standardized testing system, the IEAS, and an annual comprehensive schools survey, the ISAS. The collected data is stored, sorted, and slightly analyzed by the Assessment Entity before reports are published in the form of Schools' Articles (SAs). In addition, the Assessment Entity, and at their discretion, releases the collected data to the press. Moreover, and in quite a sporadic manner, the Assessment Entity provides the governing board with executive reports that includes key, or requested, findings of the collected data. The SAs are hence considered the main and official method to disseminate performance measurement data.

Measurement information, and its resulting potential knowledge, is not shared through emails, informal, or formal meetings across the organizational units of the case study organization. There are also no structured seminars that take place at the case study organization level to discuss and benefit form the findings of the measurement data. The only exception to this finding is the Assessment Entity where these emails, formal and informal meetings are reported to take place on regular basis.

The case study organization has not developed any policies and procedures to address the process of measurement data utilization. There are no processes in place

to facilitate or guide the use of measurement data. Employees at the Assessment Entity seem to be the only group that benefits from regular training opportunities in areas related mainly to data collection, storing, and reporting. However, there are no training opportunities, at any of the case study organizational units, in the utilization of measurement data in decision making, planning, resource allocation, and monitoring performance.

Research question number (2):

> **How was the data/information produced by performance measurement systems perceived by the recipients of this information? How was the data interpreted? For example, did it impact important issues from their point of view? Did they see it as presented in a suitable way and in a timely manner that rendered itself useful to guide or influence decision making?**

There are a number of different stakeholders that can be considered as potential recipients and users of the performance measurement data and information. The research conducts a systematic process to identify stakeholders and collects data from representatives of each of the identified groups. There are patterns of key cross-stakeholders differences in the ways measurement data are perceived.

Some stakeholder groups, such as the management cadre of the Operational Entity, entertain concerns about the quality of the measurement data in terms of the methods used to collect, store, and report on this data. These doubts are influencing their perspective of the measurement data, and ultimately their use of this data. Both this group and the governing board perceive the quality of data analysis to be particularly weak. Failure to place abstract measures in the right context, in addition to weak comparison with past years that can potentially highlight progress, is perceived to have negatively affected the position of the media and, as a result, the public at large, of the reform. Finally, constant delays in publishing measurement data have rendered them valueless to the decision making process of the management cadre group at the Operational Entity.

Other stakeholder groups, such as the media and the schools system, are suspicious of the motives behind establishing an elaborate performance measurement system to track progress in particular areas that are selected by foreign consultants. The fact that the performance measurement system was designed by western consultants, and without the participation of local stakeholders, casts quite a negative shade on the way these stakeholder groups perceive the measurement data. Moreover, these groups are largely influenced by what they think of as the leaderships' position of the performance measurement system. Perceived poor leadership support, albeit not substantiated by archival review, is further influencing their beliefs and attitudes towards measurement data.

Research question number (3):

> **How was performance measurement data/information utilized for monitoring, evaluation, reporting, decision making and strategic planning by each of the key stakeholders and how was it used differently by each of them? How was it used to influence operational decisions (staffing, building …) that reflected the intentions of these stakeholders as well as bring about changes in policies and direction (strategic, sector-related changes), and how was it used by stakeholders in a manner that was not intended by the design of the PMS?**

The findings highlight a general **underutilization of performance measurement data, by all stakeholders, in any of the practices and uses identified in the literature.** These uses include decision making, planning, evaluation, resource allocation, and monitoring. Current performance measurement data are perceived by the public officials at the case study organization as both inadequate and irrelevant to guide their policy, strategic and operational decision making process. Similarly, other stakeholder groups perceive the measurement data as irrelevant to their key concerns and needs. In addition to the overall poor use of measurement data, there is a reported misuse of data by the media stakeholder group who allegedly utilizes

select parts of the results to make a case against the reform. This 'misuse', in turn, instigates a reaction from the case study officials to defend the reform. Officials at the case study engage in a 'crisis' mode of operation where selective use of measurement data is practiced.

Moreover, the design of the reform purposely positions the utilization of performance measurement data to select schools, **parental choice**, as one of the key components of the new paradigm. This particular use of measurement data, which is not well covered in the literature, is critical to the success of the reform. Although a logistical reason related to the supply of reform schools currently prohibits parental choice, the fact that parents and students stakeholders groups are not using measurement data for this purpose should be noted and taken into consideration in a study that examines the utilization of the performance measurement system.

Research questions number (4) and (5):

> **Why were the performance measurement data/information utilized that way?**
> **What were the factors that affect and influence the process of utilization (challenges, opportunities, facilitating or inhibiting factors)?**

This research finds a **mixture of tangible and nontangible factors** that facilitate and inhibit utilization of measurement data by stakeholders. Tangible reasons are more direct and perceptible compared to the nontangible factors. They are associated with technical characteristics, organizational policies, processes, and requirements. Nontangible factors are indirect and less audible but work in profound ways to influence the perceptions, beliefs, and attitudes of stakeholders towards the performance measurement system at the case study organization. In turn, their influence is eventually affecting the decision of these stakeholders to use the measurement data that this system produces.

Tangible factors

These include certain technical characteristics of the performance measures affect the end utilization of their outcome data to support and guide decision making at all level. At the policy and strategic level, stakeholders have concerns about **the outcome focus of the measurement data**. Although this focus is by design it seems that the realities of reform implementation have highlighted new data needs that were not originally envisioned or expected. In addition, the **unsatisfactory quality of analysis** is rendering measurement data ineffectual to high level decision makers. Data is also presented out of context and with **poor comparison to data from previous years**. This makes it difficult to monitor progress in performance. Members of the governing board require quality synthesis of data, rather than semi-raw data, to support their decision making process.

At the operational level, stakeholders have concerns about the **mix and focus** of the performance measures, **frequency and tardiness** of data release, **format of files** used to disseminate data, **quality of data analysis**, and **poor comparison with past data**. Public officials who are faced with the necessity to take a myriad of operational decisions on a day to day basis find current measurement data irrelevant and difficult to use for all the listed reasons.

Additional tangible factors that influence performance data utilization include rational features borrowed from the traditional organizational theory. **Public commitment of financial resources** is well recognized in the literature to facilitate implementation and success of performance measurement systems. While the case study organization has enjoyed continuous financial backing since the inception of the reform, the potential benefits of this facilitating factor are not realized. However, negative influences caused by an evident **deficiency in technically knowledgeable staff who are devoted to professional collection and evaluation of performance measures** seem to overflow any potential benefits associated with abundant financial resources. Negative effects of the human resources factor are further amplified by clashing human resources policies at the national level.

At the organizational level, the case study has not taken actions to advance the performance measurement system and institutionalize the utilization of its outcome data. In the complete **absence of internal policies and procedures** to govern and guide implementation, the decision to utilize performance measurement data is basically left to the discretion of the employees at the case study organization. Similarly, **lack of legislative requirements** on data utilization, or otherwise externally mandated reporting requirements, means that data utilization is left to the discretionary impulse of an audience that is yet to come to terms with the new culture of measurement and overcome the many other obstacles to utilization cited in this research.

Nontangible factors

These include the perceived **implications of foreign consultants' involvement in the design of the performance measurement system**. A purposeful adoption of a unitary approach singles out schools' outcomes as the focus of measurement and overlooks the variable data needs of different stakeholders. Failure to address the complex issue of multiple stakeholders during the system's design and development stages has raised a combination of suspicion and resentment among local stakeholders.

The involvement of foreign consultants in the design of a one-sided performance measurement system is hence considered as an indirect factor that shapes the attitudes of stakeholders and ultimately contributes to their poor utilization of measurement data that does not meet their specific information needs.

In addition, two culturally rooted factors are found to be indirectly influencing utilization of measurement data to guide policy and strategic level decisions. The first factor is related to the general **absence of a measurement culture** in the country. There is no spread understanding and appreciation of the mechanisms and importance of data based decision making. Moreover, other cultural factors supersede any endeavor by public officials to take decisions based on hard data. In

particular, a cultural preference towards **pleasing the general public and face saving** takes the place of otherwise sincere attempts to utilize hard data produced by the measurement system.

Sustained, and visible, leadership support has a profound effect on the beliefs and attitudes of stakeholders towards the use of performance measurement data. Local stakeholders in particular attach great importance to what they perceive as their leadership's position of the measurement system. They purposefully align their position, and ultimately their decision to utilize the measurement data, with that of the leadership's position, as they perceive it. Several reasons are offered to explain the magnitude of attention that stakeholders assign to the public appearances and endorsements of leadership. **Public commitment of Leadership** is hence considered an indirect factor that affects performance measurement utilization.

Adopted **internal and external communication strategies and practices** contribute to the present status of measurement data utilization. The nature of the communication influence appoints it as an indirect factor. Current communication practices are not properly contributing to improving stakeholders' comprehension and appreciation of a measurement culture. The use of innovative and artfully packaged media messages to disseminate information is not sufficient to bring about the required changes in stakeholders' attitudes and practices regarding measurement data utilization. A number of communication approaches are suggested to address current gaps.

6.3   Conclusions and Recommendations

To sum up and integrate both the five key findings listed in chapter 5 and the answers to the research questions listed in this chapter, the following section draws general conclusions that cut across the different findings and answers the research questions. A number of practice and policy level recommendations are construed from these conclusions.

6.3.1 Conclusions

Despite increasing public sector interest in the topic of measurement around the world, and the consequential proliferation in research over the past two decades as illustrated in the first chapter of this research, developing performance measurement systems in the public sector continues to be a key challenge within the reform agenda (Wisniewski and Stewart, 2004). The performance measurement field continues to struggle with the question about how to assimilate measurement, and the use of measurement data, into the normal practices of public sector officials.

Moreover, any journey to explore the best ways to measure a process as complex as the production of education is indeed destined to be thorny and controversial (Propper and Wilson, 2003). Evidence abounds from around the world about the many challenges that face attempts to measure the performance of students, schools' teachers and administrators, and schools.

Furthermore, the inherent complexity of introducing major changes to an entire public sector through a reform initiative requires attention to a web of factors that extend beyond what the fields of education and performance measurement have to offer individually. Failure to simultaneously change a number of relevant interdependent factors in accordance with the nature of the selected reform is a recipe for failure and eventual abandonment of the effort (Moynihan, 2006). The adoption of public sector policy can not be at odds with an introduced reform.

The establishment of the performance measurement systems at the case study organization was pivotal to the philosophies and principles that denoted the new reform initiative. Among many other goals, the measurement system is intended to monitor the progress of reform, provide accurate data to guide decision making, and help fulfill principles of accountability and parental choice. It follows that utilization of the produced measurement data and information by concerned stakeholders is fundamental for the successful functioning of the reform. Exploring and identifying reasons why stakeholders are or are not utilizing measurement data constitute the

impetus for this research and serve as an important step towards future improvement of data utilization.

The research finds fifteen key factors to influence the implementation of the performance measurement system by facilitating or constraining the utilization of its data. The identification of these fifteen factors fills a gap in the current literature. Table 6.1 attempts to plot these factors on several fields and disciplines according to their most applicability.

6.3.1.1 A unique role for leadership visible support

The role of leadership commitment is covered quite extensively in the public reform/development literature. A number of World Bank reports highlight political commitment as one of the key factors to the success of development projects (Heaver and Israel, 1986; McCourt, 2003). At the same time, the influence of external interest groups is also recognized in the performance measurement literature. Support from elected officials (an external interest group) is reported as essential to performance measurement implementation through (1) allowing the organization to devote resources, and (2) using the information even when the results are not in breach of a political agenda (de Lancer Julnes and Holzer, 2001). However, the focus of this research's finding is on the particular role of the top leadership, versus elected officials, in convincing stakeholders to use measurement data. The field of public reform was therefore found to be more relevant to the findings of this research due to its accumulated experience in dealing with leaderships in countries undergoing reform change.

Yet this research provides new insights by revealing certain features of the culture that have an impact on the success of a performance measurement system. In a top down reform, introduced by two national leaders, citizens seem to hinge their approval, and ultimately use, of the reform's measurement data on the opinion that they believe their leadership have of the reform in general and the measurement system in particular. Utmost importance, and meaning, is attached to the public and

visible behavior of the leadership towards the reform's public events. The observed overemphasis on leadership's public appearances and behavior certainly differs from, and extends beyond, what the literature notes regarding 'offering leadership a chance to publicly express their support to a reform initiative' (Heaver and Israel, 1986). These 'soft' cultural features and behavioral attributes work in subtle ways but can eventually influence stakeholders' decision to utilize measurement data. As such, they should be taken into consideration during the design of a reform's policies and programs including adopted communication programs.

6.3.1.2 The complex needs of stakeholders

At the same time, and despite overwhelming evidence about the effect of leadership position on their beliefs and attitudes, stakeholders can not get themselves to approve, and consequently use, a system that does not account for their specific data needs. At the face of it, the two findings seem to contradict with one other. However, and while refining the understanding of the different factors affecting the utilization of measurement data and the relationship between these factors, it became apparent that factors are quite inter-related with each other and the effect of one factor can reduce, or even eliminate the impact of other factors. The actual impact of the multiple stakeholders factor overweighs many other factors, including leadership commitment. Evidence for this statement can be drawn from stakeholders' negative behavior towards measurement data during the early years of inception when leadership commitment was high and visible.

Indeed, Patton (1978) asserts that if evaluation results are to be used, the evaluation process must include the proper identification and involvement of relevant decision makers and information users. These include individuals who have questions they want to have answered (page 284). Stakeholders in the study pose a number of unanswered questions about the country's education reform. The questions, which reflect cultural and religious concerns, are not addressed by the current measurement data. Absence of attention to the specific data needs of multiple stakeholders is a main characteristic of the case study's current performance measurement system.

The system does not ensure that the interests of all stakeholders are reflected and taken into account.

It follows that due to the classically disputed nature of performance measures, a performance measurement system should employ multiple indicators that convey the different tangible and nontangible aspects of policy implementation as well as reflect the interest of all stakeholders (Van Thiel and Leeuw, 2002). Further insights into nontangible factors provided by this study include two key cultural features that represent real obstacles to measurement data utilization.

6.3.1.3 Outward-focused fear of embarrassment and accountability

Different from known 'inward-focused' concepts about an organization's 'reluctance' and 'fear of embarrassment' that can be associated with using performance measurement data (Hatry *et al*., 2003; Mayston, 1985), the study shows "outward-focused' cultural tendencies associated with the use of such data. Moreover, the dynamics of the relationship between external interest groups and the use of measurement data in this research do not mirror what is portrayed in the literature about this relationship (de Lancer Julnes, 2001). Public officials at the case study are hesitant to use measurement information, otherwise supportive of their political agenda, in fear of embarrassing local teachers, schools administrators or other groups in the community. In addition, a prevalent lack of cultural understanding and appreciation of the value and means of measurement is another key cultural feature that is not conducive to the eventual utilization of measurement data.

Perhaps spreading a measurement culture and improving the public's understanding of the values of measurement can be achieved through targeted communication strategies that approach different aspects of measurement in an integrative and stakeholders-sensitive manner. Indeed the research uncovers the need for advocacy and development communication strategies to influence audiences, support the intended change and its sustainability, and induce behavior and social change.

However, reluctance to embarrass the society and its members by publishing and using measurement data raises serious questions about a fundamental concept that underlines performance measurement: the concept of accountability (de Lancer Julnes, 2006; Moynihan, 2006). Access to information, and allowing it to circulate freely, is perceived by several researchers as a key to accountability (Holzer and Halachmi, 1996). The question then becomes how to best ensure that accountability is achieved via a performance measurement system that operates under disadvantageous cultural circumstances? What policies are most conducive to instigating social change and promoting free access of information? What processes, procedures and mechanisms need to be introduced by the case study to help institutionalize the intended changes?

6.3.1.4 Surfacing design and preparedness issues

More specific to the performance measurement system itself, there are conceivably design issues relevant to a discussion about the nontangible cultural factors identified in the research. Although the findings discussion in chapter 5 queries why none of the performance measurement frameworks used in the public sector were utilized by the case study organization, in particular to address the multiple stakeholders issue, it is doubtful that any of these frameworks takes into consideration the nontangible features identified in this research. There is a clear need for innovative research that approaches designs and frameworks of performance measurement systems in a new, and more culturally sensitive, manner.

Some of the factors that are found to influence utilization of measurement data in guiding decision making are tangible in nature and work in more direct ways. The use of measurement data as a decision guiding tool is recognized as a key value of performance measurement systems (de Lancer Julnes and Holzer, 2001). The influencing factors include quality of data analysis, comparison with past data, formats of the files used to deliver data, and frequency and timeliness of delivery schedule. These factors have been identified in the measurement literature as

features that influence the success of performance measurement systems (Holzer and Halachmi, 1996).

This discussion then raises a question about the organizational nonattendance to the importance of creating a context that accommodates and facilitates the new measurement system. There is virtually no formalized internal guidance to support and organize the expected steps that logically follow the production of data by the established measurement system. Perhaps the answer lies in the young and undeveloped performance management and measurement experience among officials at the case study organization. It could also be that the sudden nature of the change associated with the sweeping reform has resulted in unintentional neglect to learn and incorporate important lessons from the literature and the experience of other countries with performance measurement systems. The development of compatible internal policies and procedures to govern and guide utilization can clearly facilitate the implementation of a measurement system. Specific and detailed measurement 'guidelines' or 'programs' can establish the what, whom, where and when about data collection, storage, retrieval, analysis, and dissemination.

Other tangible factors borrowed from the rational model in the traditional organizational theory include committing resources and having staff devoted to evaluation of performance measures (Holzer and Halachmi, 1996; de lancer Julnes and Holzer, 2001). Once again, factors are found to be inter-related with each other. For example, the negative effect of not having knowledgeable staff committed and devoted to evaluation of performance measures is reducing, if not eliminating, the positive impact of abundantly committed financial resources.

6.3.1.5 Foreign consultants and top-down reforms

Ideas and plans presented by foreign consultants operating in foreign countries are not necessarily judged solely on their objective values; they can sometimes be adversely affected by a number of emotional factors that inherently exist in the style and pattern of interactions between consultants and consultees (Kaffman, 1961). The

reform in the case study is mainly designed and developed by a foreign consultant firm. This research provides new insights about the involvement of foreign consultants in reform initiatives in developing countries by attempting to understand the reasons behind negative local stakeholders' reactions to this involvement.

The nature of a top-down reform, and the development of the initiative and its components in isolation of local stakeholders, have contributed to the dissatisfaction, apprehension, and even suspicion of stakeholders regarding the role of foreign consultants. Moreover, education is a particularly sensitive sector given its perceived role in preserving national identity, local heritage, and religion. Fears from foreign, in particular western, ideas and philosophies were only augmented by the strong presence of western consultants. Attitudes emanating from psychological factors and fears in the background of local stakeholders play a major role in how ideas of foreign consultants, such as the performance measurement system, are perceived by these stakeholders (Kaffman, 1961).

At the same time, communication messages conveyed by the case study organization fell short of addressing stakeholders' fears and concerns regarding the role of foreign consultants. Ultimately, a number of factors that are identified in both the literature and this study seem to interrelate and play into the issue about the involvement of foreign consultants in the reform initiative and its influence on measurement data utilization.

In summary, this research provides new insights into the utilization process by identifying factors that influence this process from the perspectives of public officials and other stakeholders. It also reveals a number of implementation issues relevant to a discussion about the design and focus of a performance measurement system. The utilization dilemma surfaces as a rather convoluted notion. Its success is dependant on a mesh of tangible and nontangible factors, some of which take place during stages that well precede the actual production of data. Utilization, therefore, should not be approached in a partial mindset. Neither should it be tackled as an isolated stage or a single step in a process. Correspondingly, utilization should be

well considered right from the onset by those who wish to promote performance measurement.

6.3.1.6 The political and governmental setting

Finally, concepts and methods of performance measurement in the public sector have been evolving from pure financial control to measuring outcomes of services since the early 20th century in a country like the United States. Similarly, the use of performance information in Australia, New Zealand, and the UK serves a specific purpose and is part of the New Public Management (NPM) doctrine. NPM aims at holding managers accountable for clearly articulated goals and in turn provides them with adequate authority to achieve these goals. The progression of performance measurement in all these countries is likely consistent with the progression of their social and political systems. In fact, since the US government started emulating the UK, Australian, and New Zealand examples over the last decade, many questions have been raised regarding the compatibility of this doctrine with the institutional design of the US government (Moynihan, 2006).

However, and in all these countries, accountability has been the center notion throughout their journey with performance measurement evolution. The need for measuring the outputs and outcomes of public actions and activities represent a key element in statements and speeches made by their politicians and public officials (Van Thiel and Leeuw, 2002). Performance measurement data represent an opportunity for these governments, their public and their stakeholders to account for performance.

Perhaps it is then not perplexing to question the appropriateness and validity of concepts and theories that have been developed and tested mainly in the Anglo-American context. This calls for more than just accounting for the idiosyncrasies of developing countries or simply adapting approaches and frameworks used elsewhere. The entire set of underlying assumptions, guiding principles, and the institutional design of governments need to be examined before the concept of measuring the

performance of the public sector is even considered in different political and cultural environments.

Field	Factor	Type
Performance Measurement & Program Evaluation Literature	1. Quality of analysis	Tangible
	2. Comparison with past data	Tangible
	3. Frequency and timeliness	Tangible
	4. Format of files	Tangible
	5. Financial Resources	Tangible
	6. Committed qualified staff	Tangible
	7. External requirements	Tangible
	8. Internal policies & procedures	Tangible
Public Development Literature	9. Communication	Nontangible
	10. Leadership commitment	Nontangible
	11. Foreign consultants	Nontangible
Public Sector Management	12. Multiple stakeholders	Tangible
Education Literature	13. Outcome focus of measures	Tangible
Others	14. Absence of a measurement culture	Nontangible
	15. Pleasing the public and face saving	Nontangible

Table 6.1: Factors that Influence Utilization of Performance Measurement Data

6.3.2 Recommendations

There are clearly few problems with implementing the performance measurement system and utilizing its measurement data and information at the case study organization. It is therefore not surprising that the following recommendations focus

on schemes to improve the utilization of the performance measurement system's data and information based on the findings of this research:

■ Develop and implement a **systematic assessment process** for the performance measurement system.

An assessment process allows for unremitting attention and commitment to the development and refinement of the performance measurement system. The process must be periodic and should follow a timetable that takes into account schools' calendars, timetable of current and expected data collection activities, and other relevant events. A carefully deliberated assessment process can also contribute to addressing a recognized need for dynamic rather than static measurement systems (Neely, 2005).

The development and design of the assessment process should benefit from the findings of this research. As a minimum, all tangible factors can be used immediately as criteria or a checklist of items that need to be considered in an assessment undertaking. Addressing and measuring the identified nontangible factors will require considerably more effort to undertake and incorporate into a performance measurement system.

Moreover, the assessment process should incorporate feedback from different groups of stakeholders. Identification of stakeholders groups must follow one of the methodical methods in the literature. A vigilantly developed assessment process can improve the sense of ownership among stakeholders. Incorporating the feedback of stakeholders can also enhance the image of a measurement system currently perceived as static and not responsive to their data and information needs.

■ The second recommendation, a consequent of the first, is to **review and revise the focus, scope, and components of the measurement system.**

The revision can greatly benefit from the finding of this research as well as from evidence out of the education sector in countries that experimented with outcome based evaluation systems. Certainly the verdict on outcome based systems is yet to arrive. The literature is greatly divided between staunch proponents and unwavering opponents to the general shift towards performance measurement that focuses on outcomes. It must be added that this shift is part of a move to increasingly use performance information toward external accountability and control (Mannion and Goddard, 2000). It follows that when the use of measurement towards accountability is a contentious subject it only makes sense to avoid over-emphasis on outcomes measurement.

And while many lessons can be learned from the accumulated experience in outcome based evaluation, a fresh approach is probably what is most needed. Around the world, the use of performance assessment in the public sector is coinciding with the rise of administrative reform (Van Thiel and Leeuw, 2002). Performance measurement systems that are established within a given sector should therefore serve more than the narrow, subject-matter, data requirements of that sector. A performance measurement system for the education sector, which is created as part of reform initiative, must include aspects that reflect progress of the reform in areas outside the constricted, albeit admittedly complex, process of producing education. As it stands, there is doubt regarding the possibility of even being able to measure the multiple tasks undertaken by a school given that some of these tasks are inherently un-measurable (Dixit, 2002).

It is therefore recommended that reform and major change initiatives in the education system that consider the use of performance measurement should steer away from overemphasis on outcomes, especially subject and sector specific outcomes. This is specifically pertinent during reform commencement when the need for public officials and policy makers to closely monitor and assess progress in the reform's key inputs and processes is particularly high. It is also equally important to ensure that measures are in place to monitor the dynamics and implications of relevant policy implementation factors outside the targeted sector

such as the government's HR and labor policies, procurement and managerial flexibility. In many cases, emerging needs for remedial actions or course correction can not greatly benefit from, or most importantly wait for, outcome focused data.

◼ A third related recommendation is to **adopt a gradual and incremental approach to establishing performance measurement systems and introducing performance measures.**

This represents quite a departure from current norms and traditional practices where a full package of performance measures is pre-developed as part of establishing a performance measurement system and in advance of implementing the resulting, ready-to-use, system. The field has long been obsessed with developing the most 'perfect' frameworks to guide the development of pre-packaged measures. Moreover, the dominance of one single framework, the balanced scorecard, has probably not benefited the evolution and long term professionalization of the field (Neely, 2005). It is perhaps time to think outside the box and consider unorthodox alternatives to mainstream practices that result in the sudden drop of a fully developed, yet untested, comprehensive measurement system.

Moreover, systematic feedback from a recommended rigorous assessment process and/or research, such as action research, can contribute to the refinement, and better development, of a more efficient and relevant measurement system. While the development of the initial set of performance measures can, and probably should, benefit from the general guidance of currently available frameworks, this effort should be mainly steered by a set of policy questions and stakeholders' data needs that are bound to be unique to the context and circumstances of every single case.

The notion of 'piloting' before fully 'institutionalizing' performance measurements can prove to be a key part of the sought after answer to an identified need for flexibility in measurement systems. It can also be the method of choice in countries where measurement is a new concept and human capacity to identify data needs and develop corresponding measures is not fully developed yet.

■ The fourth recommendation is to approach questions related to a reform's **communication strategies** as key policy and management decisions that need to be addressed and settled at the early stages of the reform.

Deliberating and opting for the most fitting communication strategies should be an integrative part of efforts to address other policy questions that face the initiation of a reform. By the same token, engage and incorporate feedback from stakeholders early on in the communication development process to ensure that the reform is well understood and appreciated by the general public and have the strong support of key stakeholders. Substantial changes in the attitude and behavior of the society are required to accept, appreciate, and practice a measurement culture. Careful selection of appropriate communication strategies should be on the agenda of those who would like to promote performance measurement.

■ The fifth and final recommendation is to **view and assess the viability of new reforms, even if limited or sector-specific, in a holistic and all-inclusive manner**.

Such a view must consider the degree of harmony between the envisioned reform and the policies and activities prevalent within the larger context of the public sector. A simple example refers to the importance of ensuring that the reform policies are compatible with other public policies in the government, such as HR, labor, and procurement laws, to support the success of introduced changes.

Equally important, the view must also take into consideration the overall philosophy and system of the public sector. Compatibility of an envisioned reform with the institutional design of the government and its existing political philosophies, public administration cultures, and formal systems must be a decisive passing grade before the initiation of any reform. The establishment of market type mechanisms, the establishment of semi-autonomous public service providers, and the consequent adoption of private sector techniques such as performance indicators, are all characteristics of the New Public Management (NPM) doctrine. Holding public organizations, and consequently public officials, accountable for their performance is

an example of a principle that underpins many changes in this doctrine. In addition, inherent interconnection between NPM concepts, such as managerial flexibility and authority, means that these concepts depend on each other to work and should not be a subject for selective inclusion in a reform setting. Compatible cultures, government institutional designs, and existing formal systems may or may not be present in all countries to support these principles and interrelated concepts.

6.4   Contribution to Knowledge: Theoretical and Practical Contributions of the Research

In compliance with the requirements of scientific research, any research work must demonstrate that new knowledge has been generated. Researchers should be capable of answering the question about how their work resulted in knowing more about an issue that was unknown before. Contribution to knowledge is not only a major objective of any research work but also a determinant factor of the overall quality of case study research (Eisenhardt, 1989).

The following is a description of the contributions to theory and practice provided by this research on the topic of performance measurement data utilization. It is worth adding that given the importance of the selected public organization, and its representation of the full education sector, this case study research has an added advantage of providing evidence-based insights into the ambiguous and underdeveloped area of measurement data utilization not only in the general context of public sector reform but also in the specific context of public education reform.

6.4.1 Contribution to theory

■ *Enhanced knowledge about the factors that influence the utilization of performance measurement data*

This contribution occurred as a direct consequence of answering research questions four and five. Previous research offers isolated conclusions and gives partial answers to the overarching question about the implementation of performance measurement systems. Most of this research focuses on implementation and utilization in specific sectors such as education and health (Van Thiel and Leeuw, 2002; Propper and Wilson, 2003); in private sector settings (Bourne, 2005); in public sector organizations with focus on the issue of multiple stakeholders (de Lancer Julnes, 2001; Wisniewski and Stewart, 2004; McAdam *et al.*, 2005); in public sector organizations with distinction between policy adoption and actual implementation stages (de Lancer Julnes and Holzer, 2001; de Lancer Julnes, 2009); and in public sector setting with focus on reform initiatives and modernization (Moynihan; 2006; Sotirakou and Zeppou; 2006).

As a result of synthesizing cross-disciplinary research that cuts across a number of disciplines, this research proposes a mix of fifteen tangible and nontangible factors that influence a specific aspect of implementing performance measurement systems: the utilization of their outcome data and information. The factors also reflect stakeholder perspectives and therefore address a major research challenge in the field of performance measurement (Marr and Schiuma; 2002). Moreover, some of the factors identified in this research were not addressed by previous research and therefore make a novel contribution.

The identified factors also provide a rational foundation for future efforts to design better performance measurement systems that attend to tangible and nontangible factors and are sensitive and responsive to stakeholders' perspectives.

Accordingly, this contribution fills a gap in current research, contributes to theory, and has important implications that can be used to close a current gap that exists between theory and practice.

■ **Portrayal of the different bearings that the identified factors have on stakeholders' decision to utilize measurement data**

This contribution transpired during the refinement of the meaning and understanding of the identified factors that affect measurement data utilization. Throughout the data collection and analysis stages, meanings and associations were sought to explain why some reportedly influencing factors proved to be less powerful than others in influencing the decision of stakeholders to actually use the measurement data for a variety of purposes. It soon became evident that the surfacing list, which initially seemed to comprise of equally important factors, in fact included factors of different impacts and importance. Stakeholders clearly attach dissimilar weights to these factors in their decision to utilize performance measurement data.

A pertinent point then is the relationship between these factors and how the impact of one factor can diminish or eliminate the impact of another. A particularly significant and novel relationship is the one that exists between committed human resources and committed financial resources: the negative impact of the former cancels out the positive impact of the latter. Another unique relationship is the identified link between the involvement of foreign consultants and stakeholders' unmet performance data needs. This research also finds that the influence of an identified factor suited to play a facilitating role (i.e. leadership commitment) can not in and by itself curtail the negative influence of some tangible factors (i.e. focus of measurement, quality of analysis, and timeliness of data delivery).

▰ A better understanding of how performance measurement systems function in countries with embryonic measurement experiences.

This contribution is a direct result of selecting a case study organization that operates in a part of the world that has not witnessed a great deal of exploration into the subject of performance measurement. Research on this topic has historically been almost exclusive to the experiences and contexts of the United Kingdom, the United States, Australia and New Zealand, and as of late some OECD countries. Few isolated studies cover other countries like Singapore but mostly with a focus on the Balanced Scorecard framework (for example Kon, 2005). Although the need to test the appropriateness, applicability, and generalizability of developed theories and

234

concepts in other parts of the world has been identified for years now (Marr and Schiuma; 2002), this has not been matched by the required level of international research work. Given the current flurry of public administration and sector reforms in developing countries, and the tendency to import reform theories and concepts from developed countries, research that looks into the dynamics of performance measurement systems in developing countries seems to be a worthwhile and timely undertaking.

6.4.2 Contribution to practice

■ *Advancement in knowledge regarding utilization paradigms helps in the design of more functional performance measurement systems*

This contribution is a result of the identification of the influencing factors and the general description of the dynamics and relationships between some of these factors. This identification and description improves knowledge about measurement data utilization and hence provides a valid and plausible base for future efforts to design more practical performance measurement systems. Practitioners can improve the conditions for utilization by designing performance measurement systems that attend to tangible and nontangible factors found to influence utilization and are sensitive and responsive to stakeholders' perspectives.

■ **Facilitates the jobs of public officials who stand to benefit from the creation of contexts that are more accommodating for the utilization of performance measurement data**

This contribution is also a result of identifying factors that influence utilization of measurement data. However, of particular relevance to this contribution are factors that are found to affect the ability of public officials to use measurement data in their policy, strategic, and operational decision making processes. Since this research has the advantage of adding the perspectives of public practitioners on the subject of

utilization, it provides significant insights to other public officials in the field who are interested or involved in performance measurement systems. Enhanced ability to use measurement data in management functions like planning, monitoring, decision making and evaluation is especially advantageous to public officials who are keen on adopting modern public management practices and improving the efficiency of their work.

◢ Provides consultants active in the field of public sector reform with valid guidance on the introduction of performance measurement systems

This contribution is a result of understanding the emerging theme about the involvement of foreign consultants in the design of the reform and refining this understanding in relationship with other influencing factors. Although the findings of this research are relevant to consultants working in their native countries, they are of particular significance to consultants who work on development projects in foreign countries. In their pursuit to arrive at the most appropriate recommendations and approaches to the reform and measurement quandaries that face their government consultees, consultants can especially benefit from the political and cultural factors identified in this research even if a review of these falls outside the scope of their assignment. This research offers practical contributions to consultants who are deliberating the establishment of performance measurement by providing them with a clear indication of the areas they need to be aware of before recommending systems that otherwise operate well in other parts of the world.

Moreover, the mere identification of the inadequacies of outcome based measurement in new education reform initiatives, and the practical implications of these inadequacies, alerts consultants working in the field to issues related to this measurement orientation and fosters their willingness to seek and experiment with other alternatives.

Lastly, this contribution is also relevant to leaderships and policy makers who are considering reform initiatives and are interested in promoting performance measurement systems in their countries.

6.5   Research Quality and Limitations

For the purpose of demonstrating the validity of this research and the reliability of its process, it is important to critically asses its quality and remain vigilantly aware and forthright about its limitations. The selected methodology and research strategy for this research was discussed in Chapter 3. A full review of the data collection and analysis process was discussed in Chapter 4. Both chapters presented the exact procedures and tactics that were applied in order to ensure the quality of the research.

The following sections address the need for critical assessment of quality by summarizing how different quality components and requirements were covered during this research process.

6.5.1 Research quality

There are many ways to demonstrate that the quality and rigour of a research process was achieved through the adoption of a logical and rational research design (Yin, 1994). The most general way to assess quality is to ensure that work done meets the four tests of validity and reliability: construct validity, internal validity, external validity, and reliability. Chapter 3 discussed these four concepts in great details along with their specific implications on any research process that uses a case study research strategy. Table 6.2 includes a brief description of the ways that this research meets the tests of construct validity, internal validity, external validity, and reliability.

Feature	How it was satisfied
Construct Validity	Research design ensured that measures used were linked to the specific research problem. Data source triangulation. Key informants asked to review transcripts of their interviews before data analysis. Raw data during data collection retained by researcher
Internal Validity	Data source triangulation. Researcher self-monitoring technique during data collection and analysis. Explanation-building for uncovered patterns and mechanisms that might have produced them during data analysis
External Validity	Use of predetermined questions during research design. Thick description of data collection stage. Case study data base. Specific procedures for coding and analysis during data analysis. Continuos comparison of evidence with extant literature during data analysis to outline contribution and generalize findings to other environments (within the boundaries of the research)
Reliability	Documentation of processes and procedures. Case study protocol. Case study database. Use of carefully recorded observations during data collection and analysis. Awareness of potential personal biases as a result of the professional working relationship between the researcher and some interviewees. Observance of quality control criteria in conducting the research

Table 6.2: Summary of Validity and Reliability Features

In addition to meeting the validity and reliability tests, other features characterize work done in this research and serve as further testimony to the quality and rigour of this study. The following discussion describes these features in a chronological order whenever possible. Table 6.3 offers a brief overview of these quality features.

Prior to articulating the research questions, a critical review of the literature was undertaken to examine current knowledge mainly in the performance measurement field but also in relevant fields. A key element of this thesis has been the new lessons learned through crossing over the boundaries of several fields and disciplines: performance measurement, evaluation, education, public reform, and public sector management. The surfacing of the emergent codes later on called for further review of the literature on development projects in the public sector mainly through work done by international institutions like the World Bank. The integration of all theses new disciplines not only allowed the researcher to better understand the utilization of performance measurement data in the context of public education reform but also increased the overall quality of the research findings. The researcher was able to assure, to the extent possible, that explanations given for certain events reflect influences from other fields.

After a careful and substantiated identification of the research problem, this thesis clearly documents the process and logic that the researcher followed in order to link the research problem, and the research questions thereafter, with the findings and conclusions. The logic behind the adoption, and appropriateness, of the case study as a research method to address the research questions was clearly demonstrated as was the process used for data generation and analysis.

A careful identification of the case study's key stakeholders, rooted in a thorough search for the main actors in relevant records and documents, served as another quality measure for a research that examines an issue from the perspectives of stakeholders. Evidence from documentations was further supported by the researcher's working experience to produce an initial list of key stakeholders. All respondents in the interviews and focus group were later on asked to identify relevant stakeholders in an attempt to cross check their feedback with the initial list. The final list of seven key stakeholders groups represented a synthesis of findings from archival review, the researcher's insights, and feedback from interviewees and focus group respondents.

Furthermore, the Interest-Power grid (Ackerman and Eden, 2003) was used to help determine the relative significance of stakeholders groups. The identification of stakeholders' power levels, and their significance to the policies and programs of the case study organization, proved to be particularly relevant to the data analysis process. During endeavors to arrive at meanings and explain phenomenon, the researcher had to often times highlight, or otherwise attach more weight, to data collected from the more significant groups of stakeholders.

Triangulation can be particularly beneficial in compensating for the weaknesses inherent in each of the used data collection method and dataset by the strengths of the other used method and dataset. The researcher was quite mindful about the expected benefits of triangulation since the beginning of this research journey.

Data was collected for this research from 23 public officials and relevant respondents, 17 publications and reports, and 1009 media records using different instruments (in-depth, semi-structured interviews, focus group, and documentation review). One objective of using these multiple data sources during the data collection stage was to ensure that a full picture will be obtained about the examined questions. This enabled the researcher to later on establish a chain of evidence and consequently detect disproving and proving evidence. Another objective was to reduce the effects of biases associated with self-interest and self-protection that can be associated with personal interviews and focus groups.

The planned initiation of data analysis during the ongoing data collection effort was to assure that the researcher has the ability to collect data needed to test alternative interpretations and to make rapid adjustment in design if need be. Although the need to collect additional data or adjust the design did not come up, the concurrence of data analysis with data collection represented a technique that was deliberately taken by the researcher to increase the quality of this research.

The search for disproving-proving evidence during the data analysis stage aimed at ensuring that alternative interpretations have been thoroughly searched for and

checked. This was done through the detection of occurrences that do not fit the general pattern and trying to understand the reasons behind deviating observations and trends.

An example of particular relevance was the blaring gap between proving evidence from interviewees and respondents to the focus group and disproving evidence from archival media records regarding the phenomena of declining leadership commitment and the resulting increase in media attacks on the reform. Eventually, the research concluded that the claimed effect of leadership's public commitment on the position of the media was not substantiated in the archival review. However, a number of conclusions were drawn from the combined data collected for this research and offered as possible explanations to the perspectives of respondents on this issue. These conclusions assumed the effect of the nature of a top-down reform, the inefficiency of the reform's communication strategies, and the overall lack of clarity about the reform and its objectives. The use of pattern matching techniques during the data analysis stage allowed for the direct assessment of how convincing the evidence of the conclusions is.

Throughout the course of this study, the researcher was continuously making a deliberate effort to create a connection with existing literature on the topic. The findings regarding the tangible and nontangible factors that influence measurement data utilization were cross-examined with existing literature from multiple fields to sharpen the theory and prove the novelty of this research.

The contribution of this research to knowledge through its theoretical and practical inputs was discussed in details in this chapter under section 6.4.

Stage	Feature	How it was satisfied
Research Design	Cross-disciplinary review	Thorough critical review of literature from the fields of performance measurement, program evaluation, education, public reform and development, and public

		sector management
	Careful and systematic identification of stakeholders and their significance	Identification followed a reiterative process anchored in evidence from documentations and supported by researcher's working experience. All interviewees asked to identify list of stakeholders to cross check the researcher's list. Interest-Power grid used to help determine the relative significance of stakeholders groups
Data Collection and Analysis	Data Source Triangulation	Review of 17 official records and publications, review of 1009 media records, in-depth personal interviews with 11 public officials, and focus group with 12 respondents
	Analysis concurrent with Data Collections	Analysis started immediately after verifying interview scripts with an interviewee and while data collections efforts were still ongoing
	Search for disproving-proving evidence	Occurrences that did not fit the general pattern were highlighted. Attempts to understand reasons behind deviating trends by examining other data sources and answers to other research questions. Conscious effort to check for alternative interpretations
	Pattern matching techniques	Similar observations and occurrences were highlighted and color coded across data from different sources. Detailed notes were taken regarding detected patterns and their sources especially within and across key stakeholders groups. Drawn conclusions were subsequently assessed against identified patterns.
Reporting of the Findings	Contribution to theory	Enhanced knowledge about factors that influence data utilization Description of the impact of, and relationships between, influencing factors Improved understanding of PMS in other countries
	Contribution to	Foundation to create functional measurement

	practice	systems
		Support to public officials seeking ideal contexts for data utilization
		Guidance to consultants to design better focused measurement systems in reform settings

Table: 6.3: Quality Features of this Case Study Research

(Source: Adapted from the GAO Case Study Evaluation, November 1990)

6.5.2 Limitation of the research methodology

In addition to highlighting the quality features of this research, another part of demonstrating its quality and reliability calls for acknowledging and documenting the limitations of the methodology that was adopted in this research.

❒ The **subjective nature of the study** is a key limitation that feeds into the ongoing, and intense, debate on how to best judge the validity and quality of qualitative research. The debate even questions the value and uses of those meanings that emerge out of qualitative research. Doubts regarding the rigor of the analysis process mostly originate from the relative absence of standardized procedures that can be followed in all cases similar to those present in quantitative research.

Although quite distinct from quantitative statistical analysis both in procedures and goals, qualitative analysis is nonetheless both systematic and intensely disciplined (Frechtling and Sharp, 1997). Qualitative analysis is also different in that the focus is on the flow of the analysis to make sure it makes sense in relation to the objectives of the research and the data that was presented. Patton (1990) comments about the need of judgment and creativity in applying guidelines to qualitative research: "*Because qualitative inquiry depends, at every stage, on the skills, training, insights, and capabilities of the researcher, qualitative analysis ultimately depends on the*

analytic intellect and style of the analyst. The human factor is the greatest strength and the fundamental weakness of qualitative inquiry and analysis".

Moreover, the fact that a case study method inherently requires subjective and judgmental components should not undermine the value of the method. A number of steps, techniques, and procedural safeguards can be utilized to 'offset' any negative effects associated with these components. As discussed before, the researcher carefully designed and adopted a research process that incorporates a number of techniques suggested in the literature (e.g. use of predetermined questions during research design, thick description of data collection stage, use of specific procedures for coding and analysis during data analysis) to minimize bias.

❐ Another key limitation of this study is **lack of generalizability**; a common criticism of case studies research. In this research, generalizability is concerned with knowing whether the findings of the study are applicable beyond the immediate case study organization. It is also concerned with the domain to which the research findings can be generalized (Yin, 1994).

This limitation call for revisiting the underlying mission that this adopted research strategy was supposed to serve. The focus of the researcher was on attaining new in-depth understanding about measurement data utilization and explanations and interpretations about the rich practical experiences of the practitioners and stakeholders; experiences that were until this research mostly unknown and unexplained. It follows that despite the difficulty to generalize results of a case study, this is a qualitative research method intended to provide a detailed account of a particular case situation. Although the method has minimal claims to generalization, neither are its findings intended for generalization or representation of other organizations, it can still highlight and enrich relevant concepts and hence offer normative suggestions without precise prescriptions (Hoskisson *et al*., 1999). In addition, the findings are particularly relevant to the field of performance measurement but also to the fields of education, public sector management, and public reform and development.

❏ The last research methodology limitation is the **lack of a longitudinal viewpoint.** This limitation is a result of time constraint. The collected data represented a snapshot of the case study that captured whatever was transpiring at the point of data collection. A longitudinal research could have allowed for a better understanding of the issues around measurement data utilization and facilitated the assessment of the impact of each of the identified influencing factors.

6.5.3 Limitation of the research findings

In order to demonstrate the quality and validity of the findings of this research, it is important to acknowledge and document the limitations of these results.

❏ A key limitation of the research findings is that the definition and scope of some of the identified **influencing factors are quite general and require further scrutiny.** This is especially true for factors that originated from emerging themes. The foreign consultant involvement factor is one case in point. Moreover, the **exact nature, and weight, of the influence of all the identified factors are not clear**. Similarly, more examination into the **dynamics and relationships between these findings** is called for. For example, this research was not able to talk with any certainty to the existence of causal relationship between the factors. This limitation of the findings is associated with the exploratory and explanatory nature of the research which focused on proposing knowledge, versus validating knowledge. The area of research interest is largely underdeveloped and knowledge is just starting to get built up. This research attempts to identify and uncover the general nature of the factors that influence utilization, versus validating these factors and confirming the extent of their influence.

❏ Another limitation of the research results is that these are **based mainly on the perspectives of practitioners and stakeholders**. As important as these views are to any management research, they capture only what is on the radar screen of these stakeholders. It follows that an associated limitation is that the **adoption of the**

research results will not necessarily ensure improved utilization of performance measurement data. It is likely that other influencing factors have not been identified in this research because they are outside the interest or knowledge of stakeholders. It is also likely that these other factors, or at least a subset of them, are simply not known to us at this point of time and will only get unearthed as the body of knowledge accumulates in this area.

❏ The last limitation is that this research, by design, **concentrates mainly on utilization in a rather limited sense of the word**. The literature review pointed towards a gap that exists specifically in the area of actual use of measurement data. Even though the research questions tried to track the utilization process after the production of measurement data, there was admittedly scant, even if purposeful, attention to what transpired before the production stage. Previous sections of this thesis concluded that utilization is a complex concept that should not be tackled as an isolated stage or a single step in a process. The rather narrow, versus inclusive, inquiry into the topic of utilization could have inadvertently contributed to the limitation of the research findings.

6.6   Areas for Future research and Additional Issues Raised by the Research

With the progression of this study, it became clear to the researcher that in order to advance knowledge in the field of performance measurement there is a need for future investigation and empirical research in several key areas. These are as follows:

❏ As we become more aware of the different factors that influence utilization of performance measurement data, further research is needed to investigate and clarify the importance and influence of the different factors highlighted in this and other relevant research. This can ideally be done through longitudinal, follow up studies.

❐ Innovative research is needed to explore ways to arrive at practical and more usable performance measurement frameworks. Given the overwhelming evidence about challenges associated with using data produced by current performance measurement frameworks, it is prudent to start questioning the design and orientation of these systems. Moreover, the illustrated dominance of few authors and frameworks in the field of performance measurement is another compelling argument in favor of the need for original research in the field. Also, the need for this kind of research is further highlighted by the expansion of performance measurement concepts practices into new countries around the world.

❐ More collaboration between researchers from different countries around the world is called for to investigate performance measurement systems from an international perspective. This kind of collaboration is required to exchange learning, avoid duplication of academic effort and working in silos, and test the appropriateness and applicability of measurement concepts and methods that evolved mainly in the Western world, Australia, and New Zealand. Models to test the influence of cultural factors and governmental systems are required not only for countries with little or no past experience in performance measurement but also for countries with an established history in the field like the United States for example. Previous sections of this thesis indicated the need to develop new sets of reform notions and suppositions which can possibly change the nature of performance measurement systems as we currently know it and produce systems that are more harmonious with the different governmental systems and cultural norms around the world.

❐ This study highlights a strong need for cross disciplinary research that enables researchers and scholars to approach critical performance measurement questions from a holistic and more comprehensive viewpoint that is more appropriate to the context and setting of the public sector. Debate about the appropriateness of market driven methods, rooted in the private sector philosophy, in public sector settings has long focused on the difference between the two sectors and the issue around multiple stakeholders. However, an interdisciplinary investigation can bring together the wider perspectives of a variety of fields that have knowledge relevant to the many

247

complex issues of performance measurement in the public sector. Moreover, cross disciplinary studies can further attract the interest of a wider range of researchers and scholars which can in time contribute to the enrichment of the field.

REFERENCES

Accounts Commission for Scotland. 1998. *The measures of success: Developing a balanced scorecard to measure performance.* Edinburgh: Management Studies Unit, Accounts Commission for Scotland.

Accounts Commission for Scotland. 1999. *Measuring up to the best: A manager's guide to benchmarking.* Edinburgh: Management Studies Unit, Accounts Commission for Scotland.

Ackermann, F., & Eden, C. 2001. *Stakeholders matter: Techniques for their identification and management.* Department of Management Science Research Paper No. 2001/20, Strathclyde Business School, Glasgow.

Ackermann F., & Eden C. 2003. *Strategic management of stakeholders: Theory and practice.* Working paper, University of Strathclyde, Glasgow.

Aguilar, O. 2003. How Strategic Performance management is helping companies create business values. *Strategic Finance,* 84(7): 44-46.

Anderson, K., & McAdam, R. 2004. A critique of benchmarking and performance measurement. *Benchmarking: An International Journal,* 11(5): 465–483.

Anderson, P. F. 1986. On method in consumer research: A critical relativist perspective. *Journal of Consumer Research,* 13: 155-177.

Amis, John M. & Silk, Michael L. 2007. The philosophy and politics of quality in qualitative organizational research. *Organizational Research Methods,* 11(3): 456-480.

Atkinson, Anthony & Epstein, Marc. 2000. Measure for measure: Realizing the power of the balanced scorecard. *CMA Management,* Hamilton, 74(7): 22-28.

Attkisson, Clifford, et al. (Eds.). 1978. *Evaluation of human service programs.* New York: Academic Press.

Ballantine, J. & Brignall, S. 1996. Performance measurement in service businesses revisited. *International Journal of Service Industry Management,* 7(1): 6-31.

Benbasat, I. 1984. An analysis of research methodologies. In F. W. McFarlan (Ed.), *The information systems research challenge.* Boston: Harvard Business School Press.

Bendheim, C. & Graves, S. 1998. Determining best practice in corporate-stakeholders relations using data envelopment: An industry level study. *Business and Society,* 37(3): 306-339.

Berkowitz, S. 1996. Using qualitative and mixed method approaches. In R. Reviere, S. Berkowitz, C. C. Carter, & C. Graves-Ferguson (Eds.), *Needs assessment: A creative and practical guide for social scientists.* Washington, DC: Taylor & Francis.

Berman, E., & Wang, X. 2000. Performance measurement in US counties: Capacity for reform. *Public Administration Review,* 60(5): 409–420.

Berman, E. 2002. How Useful is Performance Measurement. **Public Performance & Management Review,** 25(4): 348-351.

Boland, T. & Fowler, A. 2000. A systems perspective of performance management in public sector organizations. *The International Journal of Public Sector Management,* 13(5): 417-446.

Bouckaert, G. 1992. Public productivity in retrospective. In M. Holzer (Ed.), *Public Productivity Handbook:* 5–46. New York: Marcel Dekker.

Bouckaert, G. & Peters, B. G. 2002. Performance measurement and management: The Achilles Heel in administrative modernization. *Public Performance & Management Review,* 25(4): 359.

Bourne, M., Mills, J., Wilcox, M. & Neely, A. 2000. Designing, implementing and updating performance measurement systems. *International Journal of Operations & Production Management,* Bradford, 20(7): 754-769.

Bourne, M., Neely, A. Platts, K. & Mills, J. 2002. The success and failure of performance measurement initiatives: Perceptions of participating managers. *International Journal of Operations & Production Management,* 22(11): 1288-1310.

Bourne, M., Neely, A., Mills, J. & Platts, K. 2003. Implementing performance measurement systems: A literature review. *International Journal of Business Performance Management,* 5(1): 1-24.

Bourne, M. 2005. Researching performance measurement Ssstem implementation: The dynamics of success and failure. *Production Planning & Control,* 16(2): 101-113.

Bowden, M. 1996. *Development and use of outcome information in government: Oregon Department of Transportation.* Washington, DC: ASPA, Center for Accountability and Performance.

Boyne, G., Martin, S. & Walker, R. 2004. Explicit reforms, implicit theories and public service improvement. *Public Management Review,* 6(2): 189-210.

Brignall, S. & Modell, S. 2000. An institutional perspective on performance measurement and management in the "New Public Sector". *Management Accounting Research,* 11(3): 281-306.

Bryson, J. M. 1995. **Strategic planning for public and non-profit organizations.** San Francisco, CA: Jossey-Bass.

Bryson, J. M. 2004. What to do when stakeholders matter; Stakeholder identification and analysis techniques. *Public Management Review,* 6(1): 21-53.

Burgess, S., Croxson, B., & Gregg, P. 2001. *The intricacies of the relationship between pay and performance for teachers: Do teachers respond to performance related pay schemes?* Working Paper no. 01/035, CMPO, University of Bristol, Bristol.

Burgess, S., Propper, C., & Wilson, D. 2002. *Does performance monitoring work? A review of the evidence from the UK public sector, excluding health care.* Working Paper no. 02/049, CMPO, University of Bristol, Bristol.

Bush, P. 2005. Strategic performance management in government: Using the balanced scorecard. *Cost Management,* 19(3): 24-31.

Cabanero-Verzosa, C., & Garcia, H. R. 2009. *Building commitment to reform through strategic communication: The five key decisions.* Washington, DC: World Bank.

Cannon, J. 1996. *Utah tomorrow.* Washington, DC: American Society for Public Administration, Center for Accountability and Performance.

Campbell, D. 1975. Degrees of freedom and the case study. *Comparative Political Studies,* 8: 178-193.

Campos, J. E. & Esfahani, H. S. 2000. Credible commitment and success with public enterprise reform. *World Development,* 28(2): 221-243.

Caplan, G. 1960. *Some problems of technical consultation in Israel.* Paper presented at the annual meeting of the American Orthopsychiatric Association, Chicago.

Caplan, G. 1964. *Principles of preventive psychiatry.* New York: Basic Books.

Caplan, G. 1970. *The theory and practice of mental health consultation.* New York: Basic Books.

Chang, L., Lin, S. & Northcott, D. 2002. The NHS performance assessment framework: A "Balanced Scorecard" approach?. *Journal of Management in Medicine,* 16(5): 345-358.

Clark, R. 2008. An Exploratory Study of Information Systems Researcher Impact. *Communications of the Association for Information Systems*, 22(1): 1-32.

Clegg, S. R. 2002. Lives in balance: A comment on Hinings and Greenwood's disconnects and consequences in organizational theory? *Administrative Science Quarterly,* 47: 428-441.

Coleman, J. S. 1966. *Equality of educational opportunity (COLEMAN) Study (EEOS), 1966.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Cooksy L. J., Gill P, & Kelly P. A. 2001. The program logic model as an integrative framework for a multi-method evaluation. *Evaluation and Program Planning,* 24(2): 119-128.

Courty, P. & Marschke, G. 2003. Dynamics of performance-measurement systems. *Oxford Review of Economic Policy,* 19(2): 268-283.

Crandon, D. S. & Merchant, K. A. 2006. Principles to guide the development and use of effective performance measures. *Performance Improvement,* 45(2): 17-23.

Cullen, J. B., & Reback, R. 2002. *Tinkering towards accolades: School gaming under a performance accountability system.* Ann Arbor: University of Michigan Press.

Cutler, T. & Waine, B. 2005. Incentivizing the Poor Relation: 'Performance' and the Pay of Public-sector 'Senior Managers'. *Competition and Change,* 9(1): 75-87.

Davies, B., Baron, S., Gear, T., & Read, M. 1999. Measuring and managing service quality. *Marketing Intelligence & Planning,* 17(1): 33–40.

Davies, I. 1999. Evaluation and performance management in government. *Evaluation*, 5(2): 150.

De Lancer J. P. 2001. Does participation increase perceptions of usefulness: An evaluation of a participatory approach to the development of performance measures. *Public Performance & Management Review,* 24(4): 403-418.

De Lancer J. P. & Holzer, M. 2001. Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation. *Public Administration Review,* 61(6): 693- 709.

De Lancer J. P. 2006. Performance measurement, an effective tool for government accountability? The debate goes on. *Evaluation,* 12(2): 219-235.

De Lancer J. P. 2009. *Performance-based management systems, effective implementation and maintenance.* Boca Raton, FL: CRC Press.

Denzin, N. K. & Lincoln, Y. S. 1994. *Handbook of quantitative research.* London: Sage Publications.

Denzin, N. K., & Lincoln, Y. S. 2005. Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* *(3rd ed.):* 1–32. Thousand Oaks, CA: Sage Publications.

Dixit, A. 2002. Incentives and organizations in the public sector: An interpretive review. *Journal of Human Resources,* 37(4): 696-727.

Eccles, Robert G. 1991. The performance measurement manifesto. *Harvard Business Review,* 69(1): 131-137.

Economic and Social Research Council. 1994. *Building partnerships: Enhancing the quality of management research (The Bain Report).* ESRC, Sweden.

Eden, C. & Ackermann F. 1996. "Horses for Courses": A stakeholder approach to the evaluation of GDSSs. *Group Decision and Negotiation,* 5(4-6): 501-519.

Eden, C. & Ackermann, F. 1998. *Making strategy: The journey of strategic management.* London: Sage Publications.

Eisenhardt, K. M. 1989a. Building theories from case-study research. *Academy of Management Review,* 14(4): 532-550.

Epstein, M. J. 2002. Measuring and managing performance in the 21st century. In Neely, C. Adams, & M. Kennerley (Eds.), *The performance prism: The scorecard for measuring and managing business success:* 1–14. London: Financial Times-Prentice Hall.

Ewing, P., & Lundahl, L. 1996. The balanced scorecards at ABB Sweden: The EVITA project. In **The International Workshop on Cost Management, 1996**.

Faucett, A., & Kleiner, B. H. 1994. New developments in performance measures of public programmes. *The International Journal of Public Sector Management,* 7(3): 63–   71.

Feit, D. 2003. Measuring performance in the public sector. *Cost Management,* Boston, 17(2): 39-45.

Fitzgerald, L. 1991. *Performance measurement in service businesses.* London: Chartered Institute of Management Accountants.

Flick, U., 1992. Triangulation revisited: Strategy of validation or alternative?. *Journal for the Theory of Social Behaviour,* 22(1): 75–198.

Fountain, J. 1997. Are State and local governments using performance measures. *PA Times,* 20(1): 8.

Frechtling, J., & Sharp, L. (Eds.). 1997. *User-friendly handbook for mixed method evaluation.* Washington, DC: Directorate for Education and Human Resources, Division of Research, Evaluation and Communication, National Science Foundation.

Gall, M. G. et al. 1966. **Educational research: An Introduction** (6th ed.). White
Plains, NY: Longman.

Gartner. 2000. "Mapping knowledge to the process: Directions to success," web
page. As of July 20, 2010:
http://www.gartner.com/DisplayDocument?id=301719&acsFlg=accessCanBuy

Gephart, R. P. 2004. From the editors: Qualitative research and the Academy of
Management Journal. *Academy of Management Journal,* 47: 454-462.

Ghalayini, A. M. & Noble, J.S. 1996. The changing basis of performance
measurement. *International Journal of Operations and Production Management,*
16(8): 63–80.

Ghalayini, A., Noble, J., & Crowe, T. 1997. An integrated dynamic performance
measurement system for improving manufacturing competitiveness. *International
Journal of Production Economics,* 48(3): 207–225.

Gianakis, G. A. 2002. The promise of public sector performance measurement:
Anodyne or Placebo?. *Public Administration Quarterly,* 26(1): 35-64.

Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott & M. Trow 1994.
*The new production of knowledge: The dynamics of science and research in
contemporaries societies.* London: Sage Publications.

Gibbs, C, and K. Sergeant. Opportunity Not Hard Work: Scripted Solutions to
Solving Our Bibliometric Nightmare. ALIA Online, Sydney, January, 2009: 20-22.
http://www.information-
online.com.au/sb_clients/iog/data/content_item_files/000001/PresentationB6.pdf

Goddard, M., Mannion, R., & Smith, P. 2000. Enhancing performance in health
care: A theoretical perspective on agency and the role of information. *Health
Economics,* 9: 95-107.

Goldstein, H. 2003. *A commentary on the secondary school value added performance tables for 2002*. http://www.ioe.ac.uk/hgpersonal/value added-commentary-jan03.htm, October 18, 2008.

Gomes, R. C., & Liddle, J. 2009. The balanced scorecard as a performance management tool for third sector organizations: the case of the Arthur Bernardes Foundation, Brazil. *Brazilian Administration Review,* 6(4): 354-366.

Gore, A. 1993. *From red tape to results: Creating a government that works better and costs less—Accompanying report of the National Performance Review*. Washington, DC: Office of the Vice President.

Greiling, D. 2005. Performance measurement in the public sector: The German experience. *International Journal of Productivity and Performance Management,* 54(7): 551-567.

Guba, E. G. & Lincoln, Y.S. 1994. Competing Paradigms in Qualitative Research. In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of Qualitative Research:* 105-117. Thousand Oaks, CA: Sage Publications.

Guba, E. G., & Lincoln, Y. S. 2005. Paradigmatic controversies, contradictions, and emerging confluence. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed.): 1–32. Thousand Oaks, CA: Sage Publications.

Halachmi, A. & Bouckaert G. (Eds.). 1996. *Organizational performance measurement in the public sector*. Westport, CN: Quorum.

Halachmi, A. 2005. Performance measurement: Test the water before you dive in. *International Review on Administrative Sciences,* 71(2): 255-266.

Harbor, J. L. 1997. *The basics of performance measurement*. Portland, OR: Productivity Press.

Harborne, R. 1999. Power planning: An integrated business planning process. *Strategic Finance,* 81(4): 47.

Harris, J. 1998. Performance models: Enhancing accountability in academe. ***Public Productivity & Management Review,*** 22(2): 135–139.

Hatry, H. P., Morley, E., Rossman, S. B., & Wholey, J. 2003. ***How federal programs use outcomes information: Opportunities for federal managers.*** Arlington, VA: IBM Endowment for the Business of Government.

Haylett, C. H., & Rapoport, L. 1964. Mental health consultation. In L. Bellak (Eds.), ***Handbook of community psychiatry and community mental health:*** 319-339. New York: Grune and Stratton.

Healy, M. & Perry, C. 2000. Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. ***Qualitative Market   Research: An International Journal,*** 3(3): 118-126.

Heaver, R. & Israel, A. 1986. ***Country commitment to development projects,   discussion.*** Working Paper no. 4, World Bank, Washington, DC.

Hepworth, P. 1998. Weighing it up: a literature review for the balanced scorecard. ***Journal of Management Development,*** 17(8): 559-563.

Hiromoto, T. 1988. Another hidden edge: Japanese management accounting. ***Harvard Business Review,*** July-August: 22-26.

Hirschman, E. C. 1986. Humanistic inquiry in marketing research: Philosophy, method and criteria. ***Journal of Marketing Research,*** 23:237-249.

Holzer, M. & Halachmi, A. 1996. Measurement as a means of accountability. ***International Journal of Public Administration,*** 19(11/12): 1921-44.

Holzer, M. & Gabrielian, V. 1998. Five Great Ideas in Public Administration. In **Handbook of Public Administration**, 2nd ed., edited by J. Rabin, W. Hildreth, and Gerald Miller, 49-101. New York: Marcel Dekker.

Holzer, M. & Kloby, K. 2005. Public performance measurement; an assessment of the state-of-the-art and models for citizen participation. *International Journal of Productivity and Performance Management,* 54(7): 517-532.

Horton, M. 2004. *The law of perfect freedom.* Chicago, IL: Moody.

Hoskisson, R. E., Hitt, M. A., Wan, W. P., & Yiu, D. 1999. Theory and research in strategic management: Swings of a pendulum. *Journal of Management,* 25(3): 417-456.

Huff, A., Tranfield, D., & van Aken, J. E. 2006. Management as a design science mindful of art and surprise: A conversation between Anne Huff, David Tranfield, and Joan Ernst van Aken. *Journal of Management Inquiry,* 15: 413.

Hughes, J. A. & Sharrock, W. W. 1997. *The philosophy of social research* (3rd ed.). Essex: Longman

Hunt, S. D. 1990. Truth in marketing theory and research. *Journal of Marketing,* 54(3): 1-15.

Hurmerinta-Peltomäki, L. & Nummela, N. 2006. Mixed methods in international business research. *Management International Review,* 46(4): 439-460.

Ishikawa, K. 1985. *What is Quality Control? The Japanese Way*. Englewood Cliffs, NJ: Prentice Hall

Jayarajah, C., & Branson, W. 1995. *Structural and sectoral adjustment: World Bank experience 1980–1992—World Bank Operations Evaluation Study.* Washington, DC: World Bank.

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.

Johnson, H. T. 1983. The search for gain in markets and forms: A review of the historical emergence of management accounting systems. *Accounting, Organizations and Society,* 8(2–3): 139–146.

Johnson, H. T. & Kaplan, R. S. 1987. *Relevance lost – The rise and fall of management accounting.* Boston, MA: Harvard Business School Press.

Johnson, H. T. 1992. *Relevance regained.* New York, NY: The Free Press.

Johnson, J. & Branson S. 1993. *Borrower ownership of adjustment programs and the political economy of reform.* Discussion paper no. 199. Washington, DC: World Bank.

Kaffman, M. 1961. Patterns of interaction between consultants and consultees in Israel. *Behavioral Science,* 6(4): 303-307.

Kane, T. J., & Staiger, D. O. 2002. The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives,* 16(4): 91-114.

Kaplan, R. D. 1984. Yesterday's accounting undermines production. *Harvard Business Review,* 62: 95-101.

Kaplan, R. S. & Norton, D. P. 1992. The balanced scorecard – Measures that drive performance. *Harvard Business Review,* January-February: 71-79.

Kaplan, R. S. & Norton, D. P. 1996a. Using the balanced scorecard as a strategic management system. *Harvard Business Review,* January-February: 75-85.

Kaplan, R. S. & Norton, D. P. 1996b. *The balanced scorecard: Translating strategy into action.* Cambridge, MA: Harvard Business School Press.

Kaplan, R. S. & Norton, D. P. 2001. Transforming the balanced scorecard from performance measurement to strategic management: Part I. *American Accounting Association,* 15(1): 87-104.

Kaplan, R. S. & Norton, D. P. 2004. The strategy map: A guide to aligning intangible assets. **Strategy and Leadership**, 32(5): 10-17.

Kennerley, M., & Neely, A. 2002. Performance measurement frameworks: A review. In Neely, A. (Eds.), ***Business performance measurement: Theory and practice.*** Cambridge: Cambridge University Press.

Killick, T. 1998. ***Aid and the political economy of policy change.*** London: Routedge.

Kon, D. 2005. Performance measurement and management: A Singapore perspective. ***Measuring Business Excellence,*** 9(3).

LeCompte, M. D. & Goetz, J. 1982. Problems of reliability and validity in ethnographic research. ***Review of Educational Research,*** 52(1): 30-60.

LeGrand, J. & Bartlett, W. (Eds.). 1993. ***Quasi-markets and social policy***. London: Macmillan.

Lincoln, Y. S. & Guba, E. G. 1985. ***Naturalistic inquiry.*** Newbury Park, CA: Sage Publications.

Ling, T. 2002. Delivering joined-up government in the UK: Dimensions, issues and problems. ***Public Administration,*** 80(4): 615–642.

Long, E. & Franklin, A. L. 2004. The paradox of implementing the government performance and results acts: Top-down direction for bottom-up implementation. ***Public Administration Review,*** 64(3): 309-319.

Lingle, J., & Schiemann, W. 1996. From balanced scorecard to strategic gauges: Is measurement worth it? ***Management Review,*** 85(3): 56–61.

Maisel, L. 1992. Performance measurement: The balanced scorecard approach. ***Journal of Cost Management,*** 6(2): 47–52.

Mannion, R., & Goddard, M. 2000. *The impact of performance measurement in the NHS, Report 3: Performance measurement systems—A cross sectoral study.* Report prepared for the Department of Health, Center for Health Economics, University of York, New York.

Marr, B., & Schiuma, G. 2002. Research challenges for corporate performance measurement: Evidence from a citation analysis. In *Proceedings from the Performance Management 2002 Conference:* pp. 355–362.

Marshall, C. & Rossman, G. B. 1989. *Designing qualitative research.* Newbury Park, CA: Sage Publications.

Marshall, M., Wray, L., Epstein, P. & Grifel, S. 1999. 21st century community focus: Better results by linking citizens, government and performance measurement. *Public Management,* 81(10): 12-19.

Mayston, D. J. 1985. Non-profit performance indicators in the public sector. *Financial Accountability & Management.* 1(1): 51-74.

McAdam, R., Reid, R., & Saulters, R. 2002. Sustaining quality in the UK public sector: Quality measurement frameworks. *The International Journal of Quality & Reliability Management,* 19(5): 581- 584.

McAdam, R., & Henderson, J. 2004. Influencing the future of TQM: Internal and external driving factors. *International Journal of Quality and Reliability Management,* 21(1): 51–71.

McAdam, R., Hazlett, S., & Casey, C. 2005. Performance management in the UK public sector. *Management,* 18(3): 256–273.

McCunn, P. 1998. The balanced scorecard: The eleventh commandment. *Management Accounting,* December: 34-36.

McCourt, W. 2003. Political commitment to reform: Civil service reform in Swaziland. *World Development,* 31(6): 1015-1031.

McLaughlin, J., & Jordan, G. 1998. Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning,* 22(1): 65–72.

Mefalopulos, P. 2008. *Development communication sourcebook: Broadening the boundaries of communication.* Washington, DC: World Bank.

Meredith, J. R. et al. 1989. Alternative research paradigms in operations. *Journal of Operations Management,* 8(4): 297-326.

Merriam, S.B. 1988. *Case study research in education: A qualitative approach.* San Francisco, CA: Jossey-Bass.

Meyer, M. W., & Gupta, V. 1994. The performance paradox. *Research in Organizational Behavior,* 16: 309–369.

Meyer, M.W. 2002. *Rethinking performance measurement.* Cambridge: University Press.

Micheli, P. & Kennerley, M. 2005. Performance measurement frameworks in public and non-profit sectors. *Production Planning and Control,* 16(2): 125-134.

Micheli, P., Mason, S., Kennerley, M., & Wilcox, M. 2005. Public sector performance: Efficiency or quality. *Measuring Business Excellence,* Bradford, 9(2): 68-74.

Miles, M. B. & Huberman, M. A. 1994. *Qualitative data analysis- An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Miller, J. G. & Vollmann, T. E. 1985. The hidden factory. *Harvard Business Review,* September-October: 142-150.

Miller, C. C., Cardinal, L. B., & Glick, W. H. 1997. Retrospective reports in organizational research: A Reexamination of recent evidence. *Academy of Management Journal,* 40(1): 189-204.

Mills, J., Platts, K., Bourne, M., & Richards, H. 2002. *Strategy and performance: Competing through competences*. Cambridge, UK: Cambridge University Press.

Morisawa, T. 2002. Building performance measurement systems with the balanced scorecard approach. **Nomura Research Institute Papers,** 45: 1–17.

Moore, M. H. 1995. *Creating public value: Strategic management in government.* Cambridge, MA: Harvard University Press.

Moullin, M. 2002. *Developing excellence in health and social care.* Buckingham: Open University Press.

Moynihan, D. P. 2006. Managing for results in state government: Evaluating a decade of reform. *Public Administration Review,* 6(1): 77-89.

Munro-Faure, L. & Munro-Faure M. 1992. *Implementing Total Quality Management*. New York, NY: The Free Press.

Neely, A., Mills, J., Plans, K., Gregory, M. & Richards, H. 1994. Realizing strategy through measurement. *International Journal of Operations and Production Management,* 14(3): 140-152.

Neely, A., Gregory, M. & Platts, K. 1995. Performance measurement system design– A literature review and research agenda. *International Journal of Operations and Production Management,* 15(4): 80-116.

Neely, A. 1998. *Measuring business performance: Why, what and how.* London: Profile Books.

Neely, A. 1999. The performance measurement revolution: Why now and what next?. *International Journal of Operations & Production Management,* 19(2): 205-228

Neely, A., Mills, J., Platts, K., Richards, H., et al. 2000. Performance measurement system design: Developing and testing a process-based approach. *International Journal of Operations & Production Management,* Bradford, 20(10): 1119-1139.

Neely, A., Adams, C., & Crowe, P. 2001. The performance prism in practice. *Measuring Business Excellence,* 5(2): 6-12.

Neely, A., Adams, C., & Kennerley, M. 2002. *The performance prism: The scorecard for measuring and managing business success.* London: Financial Times-Prentice Hall.

Neely, A. 2003. Dysfunctional performance through dysfunctional measures. *Cost Management,* 17(5): 41-47.

Neely, A. 2005. The evolution of performance measurement research: Developments in the last decade and a research agenda for the next. *International Journal of Operations & Production Management,* 25(12): 1264-1277.

Neely, A. 2006. The measures of performance search for meaningful effective performance measurement is notoriously difficult to get right, but many organizations view such frameworks as an essential tool for assessing the obligations of individual employees. *Financial Times.* June 2: 5–10.

Neely, A. 2007. The search for meaningful measures. *Management Services,* Enfield 51(2): 14-17.

Nelson, J. (Ed.) 1990. *Economic crisis and policy choice: The politics of adjustment in less developed countries.* Princeton: NJ: Princeton University Press.

Newcomb, K. 2006. The balanced decision-making approach. *Debt 3,* Chicago, 21(1): 40-41.

Norman, R. 2002. Managing through measurement or meaning lessons from experience with New Zealand's public sector performance management systems. *International Review of Administrative Sciences,* 68(4): 619-628.

Nunberg, B. 1997. **Re-thinking civil service reform: An agenda for smart government.** Washington, DC: World Bank.

Nutt, P. 2002. Making strategic choices. *Journal of Management Studies,* 39(1): 67–96.

Oliveira, J. 2001 The balanced scorecard: An integrative approach to performance evaluation. *Healthcare Financial Management,* Westchester, 55(5): 42-45.

Osborne, D. & Gaebler T. 1992. *Reinventing government: How the entrepreneurial government is transforming the public sector.* Reading, MA: Addison-Wesley.

Patton, M. Q. 1978. *Utilization-focused evaluation.* Beverly Hills, CA: Sage Publications.

Patton, M. Q. 1990. *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Perrin, B. 2007. Towards a new view of accountability. In M. L. Bemelmans-Videc, L. Jeremy, & P. Burt (Eds.), *Making accountability work: Dilemmas for evaluation and for audit,* 41–59. New Brunswick, NJ: Transaction Publishers.

Pettigrew, A. 1995. Distinguished scholar address to the Organization and Management Theory Division of the U.S. Academy of Management, Vancouver.

Pilkington, A. and Liston-Heyes, C. 1999. Is production and operations management a discipline? A citation and co-citation study. *International Journal of Operations and Production Management*, 19 (1): 7-20.

Pollanen, R. 2005. Performance measurement in municipalities: Empirical evidence in Canadian context. *The International Journal of Public Sector Management,* Bradford, 18(1): 4-24.

Pollitt, C. 2000. How do we know how good public services are? In B. G. Peters and D. J. Savoie (Eds.), *Governance in the Twenty-First Century: Revitalizing the Public Service:* 199–152. Canada: Canadian Centre for Management Development/Centre canadien de gestion.

Power, M. 2000. The audit society – Second thoughts. *International Journal of Auditing,* (4): 111-119.

Propper, C. & Wilson, D. 2003. The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy,* 19(2): 250-267.

Radin, B. 2006. *Challenging the performance movement: Accountability, complexity, and democratic values.* Washington, DC: Georgetown University Press.

Rahnema, M. 1992. Participation. In W. Sachs (Ed.), *The Development Dictionary:* 116–131. London: Zed Books.

Riege, A. M. 2003. Validity and reliability tests in case study research: A literature review with "Hands-on" applications for each research phase. *Qualitative Market Research,* 6(2): 75-86.

Rivenbark, W. C. 2006. A "MAP" approach to performance management in local government. *Government Finance Review,* 22(3): 28-34.

Rowley, T. 1997. Moving beyond dyadic ties: A net work theory of stakeholder influences. *Academy of Management Review,* 22(4): 887-911.

Rush, B., & Ogborne, A. 1991. Program logic models: Expanding their role and structure for program planning and evaluation. *Canadian Journal of Program Evaluation,* 6(2): 95–106.

Saldannha, C., & Whittle, J. 1998. *Using the logical framework for sector analysis and project design: A user's guide.* Manila: Asian Development Bank.

Samset, K. **Working party on aid evaluation: Glossary of evaluation and results based management (RBM) terms.** Paris: Organisation for Economic Co-operation and Development.

Sanderson, I. 2001. Performance management evaluation and learning in modern local governments. *Public Administration*, 79(2): 297-313.

Schneiderman, A. 2001. The First Balanced Scorecard: Analog Devices, 1986-1988. *Journal of Cost Management*, September/October

Shank, J. K., Jewett, W. G. Jr. & Branstad, P. A. 1995. The perform system: Turning strategies into results. *Journal of Strategy Management,* 4th quarter: S22.

Smith, P. C. 1995. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration,* 18(2/3): 277-310.

Sotirakou, T. & Zeppou, M. 2006. Utilising performance measurement to modernize the Greek public sector. *Management Decision,* 44(9).

Tangen, S. 2004. Performance measurement: From philosophy to practice. *International Journal of Productivity and Performance Management,* 53(8): 726-737.

Thorpe, R. & Beasley, T. 2004. The characteristics of performance measurement research; Implications and challenges. *International Journal of Productivity and Performance Measurement,* 53(4): 334-344.

Todorov, R., & Glanzel, W. 1988. Journal citation measures: A concise review. *Journal of Information Science,* 14(1): 47.

Todorov, R. & Glanzel, W. 1998. Journal citation measures: A concise review. *Journal of Information Science,* 14(1): 47-56.

Tranfield, D. & Starkey, K. 1998. The nature, social organization and promotion of
management research: Towards policy. ***British Journal of Management,*** 9:341-353.

Tsoukas, H. 1989. The validity of idiographic research explanations. ***Academy of
Management Review,*** 14(4): 551-561.

United Nations Development Programme and the Arab Fund for Economic and
Social Development. 2003. ***The 2003 Arab human development report: Building a
knowledge society.*** New York: United Nations Development Programme, Regional
Bureau for Arab States.

Uribe, B., & Horton, D. 1993. ***Monitoring and evaluating agricultural research.***
Wallingford, UK: CAB International.

U.S. General Accounting Office, Program Evaluation and Methodology Division.
1990. ***Case study evaluations.*** Washington, DC: U.S. General Accounting Office.

Van Peursem, K., Pratt, M., & Lawrence, S. 1995. Health management performance.
Accounting, Auditing & Accountability Journal, 8(5): 34–70.

Van Thiel, S. 2001. ***Quangos: Trends, causes and consequences.*** Aldershot, UK:
Ashgate Publishing.

Van Thiel, S. & Leeuw, F. L. 2002. The performance paradox in the public sector.
Public Performance & Management Review, 25(3): 267-281.

Viadero, D. 2006. Race report's influence felt 40 years later: Legacy of Coleman
Study was new view of equity. ***Education Week,*** 25(41): 1.

Walsh, K. 1995. Quality through markets: The new public service management. In
A. Wilkinson and H. Willmott (Eds.), ***Making Quality Critical: New perspectives
on organisational change,*** 82-104. London: Routledge.

White, L. R. & Wood, F. 2005. Strategic, cost, and performance management in the
Coast Guard and Federal Government. *Cost Management,* Boston, 19(3): 7-
15.

Wholey J. S., Hatry H. P., & Newcomer K. E. (Eds.). 1994. *Handbook of practical
program evaluation.* San Francisco: Jossey-Bass.

Wilkins, K. 1996. *Performance measurement: A work in progress. Minnesota's
Department of Labor and Industry.* Washington, D.C.: ASPA, Center for
Accountability and Performance.

Williams, D. W. 2003. Measuring government in the early twentieth century. *Public
Administration Review,* 63(6): 643-59.

Williams, D. W. 2004. *Measurement and performance budgeting in the United
States in the 1950s and 1960s.* Paper presented at the Conference of European
Group of Public Administrators, Ljubljana.

Williamson, J. 1994. *The political economy of policy reform.* Washington, DC:
Institute for International Economics.

Wisniewski, M., & Shafti, F. 2003. **The teaching and learning of performance
measurement in UK undergraduate MS/OR degrees.** Glasgow: Department of
Management Science, University of Strathclyde.

Wisniewski, M. & Stewart, D. 2004. Performance measurement for stakeholders: The
case of Scottish local authorities. *International Journal of Public Sector
Management,* 17(3): 222-233.

WK Kellogg Foundation. 2004. *Logic model development guide.* Battle Creek, MI:
WK Kellogg Foundation.

World Bank. 2005. *Expanding opportunities and building competencies for young
people: A new agenda for secondary education.* Washington, DC: The
International Bank for Reconstruction and Development/The World Bank.

World Bank. 2008. *The road not traveled: Education reform in the Middle East and Africa.* Washington, DC: The International Bank for Reconstruction and Development/The World Bank.

World Bank Independent Evaluation Group. *From schooling access to learning outcomes: An unfinished agenda.* Washington, DC: The International Bank for Reconstruction and Development/The World Bank.

World Education Forum. 2000. *Dakar framework for action, education for all: Meeting our collective commitments.* Text adopted by the World Education Forum, Dakar, Senegal.

Yin, R. K. 1981. The case study crisis: Some answers. *Administrative Science Quarterly,* 26: 58-65.

Yin, R. K. (Eds.). 1994. *Case study research – Design and methods.* Applied Social Research Methods Series, 5. Newbury Park, CA: Sage Publications.

Yin, R. K. 2003. *Case Study Research: Design and Methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

APPENDIX 1

LIST OF CONSULTED ARCHIVAL RECORDS

Al-Rayah Newspaper, September 13th, 2004

Al-Rayah Newspaper, September 15th, 2004

Al-Rayah Newspaper, September 19th, 2004

Al-Rayah Newspaper, September 20th, 2004

Al-Rayah Newspaper, September 26th, 2004

Al-Rayah Newspaper, September 28th, 2004

Al-Rayah Newspaper, September 29th, 2004

Al-Rayah Newspaper, October 12th, 2004

Al-Rayah Newspaper, October 21st, 2004

Al-Rayah Newspaper, November 21st, 2004

Al-Rayah Newspaper, December 14th, 2004

Al-Rayah Newspaper, December 16th, 2004

Al-Rayah Newspaper, January 9th, 2005

Al-Rayah Newspaper, February 27th, 2005

Al-Rayah Newspaper, March 8th, 2005

Al-Rayah Newspaper, March 16th, 2005

Al-Rayah Newspaper, March 16th, 2005

Al-Rayah Newspaper, April 10th, 2005

Al-Rayah Newspaper, April 12th, 2005

Al-Rayah Newspaper, April 17th, 2005

Al-Rayah Newspaper, April 29th, 2005

Al-Rayah Newspaper, May 2nd, 2005

Al-Rayah Newspaper, May 10th, 2005

Al-Rayah Newspaper, May 22nd, 2005

Al-Rayah Newspaper, May 22nd, 2005

Al-Rayah Newspaper, May 24th, 2005

Al-Rayah Newspaper, May 29th, 2005

Al-Rayah Newspaper, June 12th, 2005

Al-Rayah Newspaper, June 20th, 2005

Al-Rayah Newspaper, June 30th, 2005

Al-Rayah Newspaper, June 3rd, 2007

Al-Rayah Newspaper, June 3rd, 2007

Al-Rayah Newspaper, June 3rd, 2007

Al-Rayah Newspaper, June 4th, 2007

Al-Rayah Newspaper, June 5th, 2007

Al-Rayah Newspaper, June 5th, 2007

Al-Rayah Newspaper, June 6th, 2007

Al-Rayah Newspaper, June 6th, 2007

Al-Rayah Newspaper, June 6th, 2007

Al-Rayah Newspaper, June 6th, 2007

Al-Rayah Newspaper, June 6th, 2007

Al-Rayah Newspaper, June 7th, 2007

Al-Rayah Newspaper, June 7th, 2007

Al-Rayah Newspaper, June 7th, 2007

Al-Rayah Newspaper, June 7th, 2007

Al-Rayah Newspaper, June 10th, 2007

Al-Rayah Newspaper, June 10th, 2007

Al-Rayah Newspaper, June 10th, 2007

Al-Rayah Newspaper, June 11th, 2007

Al-Rayah Newspaper, June 12th, 2007

Al-Rayah Newspaper, June 12th, 2007

Al-Rayah Newspaper, June 17th, 2007

Al-Rayah Newspaper, June 19th, 2007

Al-Rayah Newspaper, June 19th, 2007

Al-Rayah Newspaper, June 20th, 2007

Al-Rayah Newspaper, June 25th, 2007

Al-Rayah Newspaper, June 25th, 2007

Al-Rayah Newspaper, June 26th, 2007

Al-Rayah Newspaper, June 27th, 2007

Al-Rayah Newspaper, June 27th, 2007

Al-Rayah Newspaper, June 28th, 2007

Al-Rayah Newspaper, June 28th, 2007

Al-Rayah Newspaper, August 5th, 2007

Al-Rayah Newspaper, August 8th, 2007

Al-Rayah Newspaper, August 12th, 2007

Al-Rayah Newspaper, August 12th, 2007

Al-Rayah Newspaper, August 13th, 2007

Al-Rayah Newspaper, August 14th, 2007

Al-Rayah Newspaper, August 22nd, 2007

Al-Rayah Newspaper, August 22nd, 2007

Al-Rayah Newspaper, August 23rd, 2007

Al-Rayah Newspaper, August 30th, 2007

Al-Rayah Newspaper, September 3rd, 2007

Al-Rayah Newspaper, September 4th, 2007

Al-Rayah Newspaper, September 5th, 2007

Al-Rayah Newspaper, September 6th, 2007

Al-Rayah Newspaper, September 9th, 2007

Al-Rayah Newspaper, September 10th, 2007

Al-Rayah Newspaper, September 10th, 2007

Al-Rayah Newspaper, September 11th, 2007

Al-Rayah Newspaper, September 13th, 2007

Al-Rayah Newspaper, September 17th, 2007

Al-Rayah Newspaper, September 17th, 2007

Al-Rayah Newspaper, September 18th, 2007

Al-Rayah Newspaper, September 20th, 2007

Al-Rayah Newspaper, September 22nd, 2007

Al-Rayah Newspaper, September 23rd, 2007

Al-Rayah Newspaper, September 27th, 2007

Al-Rayah Newspaper, September 30th, 2007

Al-Rayah Newspaper, October 2nd, 2007

Al-Rayah Newspaper, October 3rd, 2007

Al-Rayah Newspaper, October 4th, 2007

Al-Rayah Newspaper, October 4th, 2007

Al-Rayah Newspaper, October 7th, 2007

Al-Rayah Newspaper, October 7th, 2007

Al-Rayah Newspaper, October 7th, 2007

Al-Rayah Newspaper, October 9th, 2007

Al-Rayah Newspaper, October 21st, 2007

Al-Rayah Newspaper, October 21st, 2007

Al-Rayah Newspaper, October 21st, 2007

Al-Rayah Newspaper, October 21st, 2007

Al-Rayah Newspaper, October 24th, 2007

Al-Rayah Newspaper, October 25th, 2007

Al-Rayah Newspaper, October 28th, 2007

Al-Rayah Newspaper, October 28th, 2007

Al-Rayah Newspaper, October 29th, 2007

Al-Rayah Newspaper, October 29th, 2007

Al-Rayah Newspaper, October 29th, 2007

Al-Rayah Newspaper, October 29th, 2007

Al-Rayah Newspaper, November 1st, 2007

Al-Rayah Newspaper, November 4th, 2007

Al-Rayah Newspaper, November 4th, 2007

Al-Rayah Newspaper, November 4th, 2007

Al-Rayah Newspaper, November 5th, 2007

Al-Rayah Newspaper, November 5th, 2007

Al-Rayah Newspaper, November 6th, 2007

Al-Rayah Newspaper, November 8th, 2007

Al-Rayah Newspaper, November 11th, 2007

Al-Rayah Newspaper, November 11th, 2007

Al-Rayah Newspaper, November 11th, 2007

Al-Rayah Newspaper, November 12th, 2007

Al-Rayah Newspaper, November 13th, 2007

Al-Rayah Newspaper, November 14th, 2007

Al-Rayah Newspaper, November 18th, 2007

Al-Rayah Newspaper, November 18th, 2007

Al-Rayah Newspaper, November 21st, 2007

Al-Rayah Newspaper, November 24th, 2007

Al-Rayah Newspaper, November 25th, 2007

Al-Rayah Newspaper, November 25th, 2007

Al-Rayah Newspaper, November 25th, 2007

Al-Rayah Newspaper, November 26th, 2007

Al-Rayah Newspaper, November 27th, 2007

Al-Rayah Newspaper, November 27th, 2007

Al-Rayah Newspaper, November 27th, 2007

Al-Rayah Newspaper, November 28th, 2007

Al-Rayah Newspaper, November 28th, 2007

Al-Rayah Newspaper, November 28th, 2007

Al-Rayah Newspaper, November 29th, 2007

Al-Rayah Newspaper, December 4th, 2007

Al-Rayah Newspaper, December 5th, 2007

Al-Rayah Newspaper, December 6th, 2007

Al-Rayah Newspaper, December 6th, 2007

Al-Rayah Newspaper, December 10th, 2007

Al-Rayah Newspaper, December 11th, 2007

Al-Rayah Newspaper, December 12th, 2007

Al-Rayah Newspaper, December 13th, 2007

Al-Rayah Newspaper, December 13th, 2007

Al-Rayah Newspaper, December 17th, 2007

Al-Rayah Newspaper, December 26th, 2007

Al-Rayah Newspaper, December 27th, 2007

Al-Rayah Newspaper, December 30th, 2007

Al-Rayah Newspaper, December 30th, 2007

Al-Rayah Newspaper, December 31st, 2007

Al-Rayah Newspaper, December 31st, 2007

Al-Rayah Newspaper, December 31st, 2007

Al-Rayah Newspaper, January 1st, 2008

Al-Rayah Newspaper, January 2nd, 2008

Al-Rayah Newspaper, January 2nd, 2008

Al-Rayah Newspaper, January 6th, 2008

Al-Rayah Newspaper, January 6th, 2008

Al-Rayah Newspaper, January 6th, 2008

Al-Rayah Newspaper, January 8th, 2008

Al-Rayah Newspaper, January 9th, 2008

Al-Rayah Newspaper, January 9th, 2008

Al-Rayah Newspaper, January 10th, 2008

Al-Rayah Newspaper, January 13th, 2008

Al-Rayah Newspaper, January 15th, 2008

Al-Rayah Newspaper, January 17th, 2008

Al-Rayah Newspaper, January 20th, 2008

Al-Rayah Newspaper, January 21st, 2008

Al-Rayah Newspaper, January 28th, 2008

Al-Rayah Newspaper, January 29th, 2008

Al-Rayah Newspaper, January 30th, 2008

Al-Rayah Newspaper, January 30th, 2008

Al-Rayah Newspaper, January 31st, 2008

Al-Rayah Newspaper, January 31st, 2008

Al-Rayah Newspaper, February 2nd, 2008

Al-Rayah Newspaper, February 3rd, 2008, page 21

Al-Rayah Newspaper, February 3rd, 2008, page 5

Al-Rayah Newspaper, February 4th, 2008

Al-Rayah Newspaper, February 4th, 2008, page 7

Al-Rayah Newspaper, February 4th, 2008, page 22

Al-Rayah Newspaper, February 5th, 2008

Al-Rayah Newspaper, February 13th, 2008

Al-Rayah Newspaper, February 13th, 2008

Al-Rayah Newspaper, February 13th, 2008

Al-Rayah Newspaper, February 14th, 2008

Al-Rayah Newspaper, February 18th, 2008

Al-Rayah Newspaper, February 19th, 2008

Al-Rayah Newspaper, February 19th, 2008

Al-Rayah Newspaper, February 19th, 2008

Al-Rayah Newspaper, February 19th, 2008

Al-Rayah Newspaper, February 19th, 2008

Al-Rayah Newspaper, February 21st, 2008

Al-Rayah Newspaper, February 24th, 2008

Al-Rayah Newspaper, February 26th, 2008

Al-Rayah Newspaper, February 27th, 2008

Al-Rayah Newspaper, February 27th, 2008

Al-Rayah Newspaper, February 27th, 2008

Al-Rayah Newspaper, February 28th, 2008

Al-Rayah Newspaper, February 28th, 2008

Al-Rayah Newspaper, February 28th, 2008

Al-Rayah Newspaper, February 28th, 2008

Al-Rayah Newspaper, March 2nd, 2008

Al-Rayah Newspaper, March 2nd, 2008

Al-Rayah Newspaper, March 3rd, 2008

Al-Rayah Newspaper, March 3rd, 2008

Al-Rayah Newspaper, March 3rd, 2008

Al-Rayah Newspaper, March 4th, 2008

Al-Rayah Newspaper, March 4th, 2008

Al-Rayah Newspaper, March 5th, 2008

Al-Rayah Newspaper, March 6th, 2008

Al-Rayah Newspaper, March 6th, 2008

Al-Rayah Newspaper, March 9th, 2008

Al-Rayah Newspaper, March 9th, 2008

Al-Rayah Newspaper, March 10th, 2008

Al-Rayah Newspaper, March 11th, 2008

Al-Rayah Newspaper, March 11th, 2008

Al-Rayah Newspaper, March 11th, 2008

Al-Rayah Newspaper, March 11th, 2008

Al-Rayah Newspaper, March 12th, 2008

Al-Rayah Newspaper, March 13th, 2008

Al-Rayah Newspaper, March 13th, 2008

Al-Rayah Newspaper, March 13th, 2008

Al-Rayah Newspaper, March 13th, 2008

Al-Rayah Newspaper, March 17th, 2008

Al-Rayah Newspaper, March 17th, 2008

Al-Rayah Newspaper, March 18th, 2008

Al-Rayah Newspaper, March 18th, 2008

Al-Rayah Newspaper, March 19th, 2008

Al-Rayah Newspaper, March 19th, 2008

Al-Rayah Newspaper, March 19th, 2008

Al-Rayah Newspaper, March 20th, 2008

Al-Rayah Newspaper, March 20th, 2008

Al-Rayah Newspaper, March 24th, 2008

Al-Rayah Newspaper, March 25th, 2008

Al-Rayah Newspaper, March 26th, 2008, Al-Rayah Youth

Al-Rayah Newspaper, March 30th, 2008

Al-Rayah Newspaper, March 30th, 2008

Al-Rayah Newspaper, March 30th, 2008

Al-Rayah Newspaper, March 30th, 2008

Al-Rayah Newspaper, March 30th, 2008

Al-Rayah Newspaper, March 31st, 2008

Al-Rayah Newspaper, April 1st, 2008

Al-Rayah Newspaper, April 1st, 2008

Al-Rayah Newspaper, April 1st, 2008

Al-Rayah Newspaper, April 1st, 2008

Al-Rayah Newspaper, April 3rd, 2008

Al-Rayah Newspaper, April 6th, 2008

Al-Rayah Newspaper, April 6th, 2008

Al-Rayah Newspaper, April 7th, 2008

Al-Rayah Newspaper, April 7th, 2008

Al-Rayah Newspaper, April 7th, 2008

Al-Rayah Newspaper, April 8th, 2008

Al-Rayah Newspaper, April 9th, 2008

Al-Rayah Newspaper, April 14th, 2008

Al-Rayah Newspaper, April 14th, 2008

Al-Rayah Newspaper, April 15th, 2008

Al-Rayah Newspaper, April 16th, 2008

Al-Rayah Newspaper, April 17th, 2008

Al-Rayah Newspaper, April 17th, 2008

Al-Rayah Newspaper, April 17th, 2008

Al-Rayah Newspaper, April 17th, 2008

Al-Rayah Newspaper, April 20th, 2008

Al-Rayah Newspaper, April 21st, 2008

Al-Rayah Newspaper, April 27th, 2008

Al-Rayah Newspaper, April 28th, 2008

Al-Rayah Newspaper, April 29th, 2008

Al-Rayah Newspaper, April 29th, 2008

Al-Rayah Newspaper, April 29th, 2008

Al-Rayah Newspaper, April 30th, 2008

Al-Rayah Newspaper, May 1st, 2008

Al-Rayah Newspaper, May 1st, 2008

Al-Rayah Newspaper, May 5th, 2008

Al-Rayah Newspaper, May 5th, 2008

Al-Rayah Newspaper, May 5th, 2008

Al-Rayah Newspaper, May 6th, 2008

Al-Rayah Newspaper, May 6th, 2008

Al-Rayah Newspaper, May 7th, 2008

Al-Rayah Newspaper, May 8th, 2008

Al-Rayah Newspaper, May 8th, 2008

Al-Rayah Newspaper, May 8th, 2008

Al-Rayah Newspaper, May 11th, 2008

Al-Rayah Newspaper, May 11th, 2008

Al-Rayah Newspaper, May 12th, 2008

Al-Rayah Newspaper, May 12th, 2008

Al-Rayah Newspaper, May 12th, 2008

Al-Rayah Newspaper, May 12th, 2008

Al-Rayah Newspaper, May 13th, 2008

Al-Rayah Newspaper, May 14th, 2008

Al-Rayah Newspaper, May 14th, 2008

Al-Rayah Newspaper, May 18th, 2008

Al-Rayah Newspaper, May 18th, 2008

Al-Rayah Newspaper, May 21st, 2008

Al-Rayah Newspaper, May 21st, 2008

Al-Rayah Newspaper, May 25th, 2008

Al-Rayah Newspaper, May 25th, 2008

Al-Rayah Newspaper, May 26th, 2008

Al-Rayah Newspaper, May 27th, 2008

Al-Rayah Newspaper, May 27th, 2008

Al-Rayah Newspaper, May 28th, 2008

Al-Rayah Newspaper, May 29th, 2008

Al-Rayah Newspaper, May 29th, 2008

Al-Sharq Newspaper, September 8th, 2004

Al-Sharq Newspaper, September 20th, 2004

Al-Sharq Newspaper, September 28th, 2004

Al-Sharq Newspaper, September 30th, 2004

Al-Sharq Newspaper, October 7th, 2004

Al-Sharq Newspaper, October 7th, 2004

Al-Sharq Newspaper, October 12th, 2004

Al-Sharq Newspaper, October 25th, 2004

Al-Sharq Newspaper, October 26th, 2004

Al-Sharq Newspaper, November 21st, 2004

Al-Sharq Newspaper, November 28th, 2004

Al-Sharq Newspaper, December 5th, 2004

Al-Sharq Newspaper, December 14th, 2004

Al-Sharq Newspaper, December 19th, 2004

Al-Sharq Newspaper, December 22nd, 2004

Al-Sharq Newspaper, January 2nd, 2005

Al-Sharq Newspaper, January 12th, 2005

Al-Sharq Newspaper, March 1st, 2005

Al-Sharq Newspaper, March 2nd, 2005

Al-Sharq Newspaper, March 8th, 2005

Al-Sharq Newspaper, March 8th, 2005

Al-Sharq Newspaper, March 10th, 2005

Al-Sharq Newspaper, March 10th, 2005

Al-Sharq Newspaper, March 16th, 2005

Al-Sharq Newspaper, March 16th, 2005

Al-Sharq Newspaper, March 21st, 2005

Al-Sharq Newspaper, April 29th, 2005

Al-Sharq Newspaper, May 2nd, 2005

Al-Sharq Newspaper, May 10th, 2005

Al-Sharq Newspaper, May 11th, 2005

Al-Sharq Newspaper, May 12th, 2005

Al-Sharq Newspaper, May 15th, 2005

Al-Sharq Newspaper, May 16th, 2005

Al-Sharq Newspaper, May 18th, 2005

Al-Sharq Newspaper, May 19th, 2005

Al-Sharq Newspaper, May 24th, 2005

Al-Sharq Newspaper, May 28th, 2005

Al-Sharq Newspaper, May 28th, 2005

Al-Sharq Newspaper, May 30th, 2005

Al-Sharq Newspaper, May 30th, 2005

Al-Sharq Newspaper, May 31st, 2005

Al-Sharq Newspaper, June 2nd, 2005

Al-Sharq Newspaper, June 4th, 2005

Al-Sharq Newspaper, June 5th, 2005

Al-Sharq Newspaper, June 8th, 2005

Al-Sharq Newspaper, June 12th, 2005

Al-Sharq Newspaper, June 15th, 2005

Al-Sharq Newspaper, June 15th, 2005

Al-Sharq Newspaper, June 20th, 2005

Al-Sharq Newspaper, June 21st, 2005

Al-Sharq Newspaper, June 21st, 2005

Al-Sharq Newspaper, June 2nd, 2007

Al-Sharq Newspaper, June 4th, 2007

Al-Sharq Newspaper, June 5th, 2007

Al-Sharq Newspaper, June 10th, 2007

Al-Sharq Newspaper, June 10th, 2007

Al-Sharq Newspaper, June 10th, 2007

Al-Sharq Newspaper, June 10th, 2007

Al-Sharq Newspaper, June 10th, 2007

Al-Sharq Newspaper, June 11th, 2007

Al-Sharq Newspaper, June 11th, 2007

Al-Sharq Newspaper, June 12th, 2007

Al-Sharq Newspaper, June 12th, 2007

Al-Sharq Newspaper, June 13th, 2007

Al-Sharq Newspaper, June 19th, 2007

Al-Sharq Newspaper, June 19th, 2007

Al-Sharq Newspaper, June 19th, 2007

Al-Sharq Newspaper, June 19th, 2007

Al-Sharq Newspaper, June 24th, 2007

Al-Sharq Newspaper, June 25th, 2007

Al-Sharq Newspaper, June 25th, 2007

Al-Sharq Newspaper, June 27th, 2007

Al-Sharq Newspaper, June 28th, 2007

Al-Sharq Newspaper, August 8th, 2007

Al-Sharq Newspaper, August 13th, 2007

Al-Sharq Newspaper, August 28th, 2007

Al-Sharq Newspaper, September 3rd, 2007

Al-Sharq Newspaper, September 3rd, 2007

Al-Sharq Newspaper, September 4th, 2007

Al-Sharq Newspaper, September 4th, 2007

Al-Sharq Newspaper, September 5th, 2007

Al-Sharq Newspaper, September 6th, 2007

Al-Sharq Newspaper, September 9th, 2007

Al-Sharq Newspaper, September 10th, 2007

Al-Sharq Newspaper, September 11th, 2007

Al-Sharq Newspaper, September 11th, 2007

Al-Sharq Newspaper, September 12th, 2007

Al-Sharq Newspaper, September 17th, 2007

Al-Sharq Newspaper, September 17th, 2007

Al-Sharq Newspaper, September 18th, 2007

Al-Sharq Newspaper, September 18th, 2007

Al-Sharq Newspaper, September 19th, 2007

Al-Sharq Newspaper, September 19th, 2007

Al-Sharq Newspaper, September 19th, 2007

Al-Sharq Newspaper, September 20th, 2007

Al-Sharq Newspaper, September 20th, 2007

Al-Sharq Newspaper, September 22nd, 2007

Al-Sharq Newspaper, September 22nd, 2007

Al-Sharq Newspaper, September 23rd, 2007

Al-Sharq Newspaper, September 25th, 2007

Al-Sharq Newspaper, September 25th, 2007

Al-Sharq Newspaper, September 27th, 2007

Al-Sharq Newspaper, September 27th, 2007

Al-Sharq Newspaper, September 29th, 2007

Al-Sharq Newspaper, September 30th, 2007

Al-Sharq Newspaper, October 1st, 2007

Al-Sharq Newspaper, October 1st, 2007

Al-Sharq Newspaper, October 1st, 2007

Al-Sharq Newspaper, October 2nd, 2007

Al-Sharq Newspaper, October 3rd, 2007

Al-Sharq Newspaper, October 4th, 2007

Al-Sharq Newspaper, October 4th, 2007

Al-Sharq Newspaper, October 8th, 2007

Al-Sharq Newspaper, October 8th, 2007

Al-Sharq Newspaper, October 9th, 2007

Al-Sharq Newspaper, October 22nd, 2007

Al-Sharq Newspaper, October 24th, 2007

Al-Sharq Newspaper, October 24th, 2007

Al-Sharq Newspaper, October 25th, 2007

Al-Sharq Newspaper, October 28th, 2007

Al-Sharq Newspaper, October 28th, 2007

Al-Sharq Newspaper, October 28th, 2007

Al-Sharq Newspaper, October 29th, 2007

Al-Sharq Newspaper, October 29th, 2007

Al-Sharq Newspaper, October 29th, 2007

Al-Sharq Newspaper, October 30th, 2007

Al-Sharq Newspaper, November 1st, 2007

Al-Sharq Newspaper, November 1st, 2007

Al-Sharq Newspaper, November 1st, 2007

Al-Sharq Newspaper, November 2nd, 2007

Al-Sharq Newspaper, November 5th, 2007

Al-Sharq Newspaper, November 5th, 2007

Al-Sharq Newspaper, November 6th, 2007

Al-Sharq Newspaper, November 6th, 2007

Al-Sharq Newspaper, November 7th, 2007

Al-Sharq Newspaper, November 7th, 2007

Al-Sharq Newspaper, November 8th, 2007

Al-Sharq Newspaper, November 8th, 2007

Al-Sharq Newspaper, November 8th, 2007

Al-Sharq Newspaper, November 11th, 2007

Al-Sharq Newspaper, November 11th, 2007

Al-Sharq Newspaper, November 11th, 2007

Al-Sharq Newspaper, November 11th, 2007

Al-Sharq Newspaper, November 11th, 2007

Al-Sharq Newspaper, November 12th, 2007

Al-Sharq Newspaper, November 12th, 2007

Al-Sharq Newspaper, November 12th, 2007

Al-Sharq Newspaper, November 18th, 2007

Al-Sharq Newspaper, November 18th, 2007

Al-Sharq Newspaper, November 19th, 2007

Al-Sharq Newspaper, November 19th, 2007

Al-Sharq Newspaper, November 20th, 2007

Al-Sharq Newspaper, November 22nd, 2007

Al-Sharq Newspaper, November 25th, 2007

Al-Sharq Newspaper, November 25th, 2007

Al-Sharq Newspaper, November 25th, 2007

Al-Sharq Newspaper, November 27th, 2007

Al-Sharq Newspaper, November 27th, 2007

Al-Sharq Newspaper, November 29th, 2007

Al-Sharq Newspaper, November 29th, 2007

Al-Sharq Newspaper, November 29th, 2007

Al-Sharq Newspaper, December 2nd, 2007

Al-Sharq Newspaper, December 2nd, 2007

Al-Sharq Newspaper, December 2nd, 2007

Al-Sharq Newspaper, December 2nd, 2007

Al-Sharq Newspaper, December 4th, 2007

Al-Sharq Newspaper, December 6th, 2007

Al-Sharq Newspaper, December 9th, 2007

Al-Sharq Newspaper, December 9th, 2007

Al-Sharq Newspaper, December 10th, 2007

Al-Sharq Newspaper, December 10th, 2007

Al-Sharq Newspaper, December 11th, 2007

Al-Sharq Newspaper, December 13th, 2007

Al-Sharq Newspaper, December 16th, 2007

Al-Sharq Newspaper, December 16th, 2007

Al-Sharq Newspaper, December 26th, 2007

Al-Sharq Newspaper, December 31st, 2007

Al-Sharq Newspaper, December 31st, 2007

Al-Sharq Newspaper, December 31st, 2007

Al-Sharq Newspaper, December 31st, 2007

Al-Sharq Newspaper, January 1st, 2008

Al-Sharq Newspaper, January 1st, 2008

Al-Sharq Newspaper, January 1st, 2008

Al-Sharq Newspaper, January 2nd, 2008

Al-Sharq Newspaper, January 2nd, 2008

Al-Sharq Newspaper, January 3rd, 2008

Al-Sharq Newspaper, January 6th, 2008

Al-Sharq Newspaper, January 6th, 2008

Al-Sharq Newspaper, January 7th, 2008

Al-Sharq Newspaper, January 7th, 2008

Al-Sharq Newspaper, January 7th, 2008

Al-Sharq Newspaper, January 8th, 2008

Al-Sharq Newspaper, January 9th, 2008

Al-Sharq Newspaper, January 10th, 2008

Al-Sharq Newspaper, January 10th, 2008

Al-Sharq Newspaper, January 13th, 2008

Al-Sharq Newspaper, January 14th, 2008

Al-Sharq Newspaper, January 15th, 2008

Al-Sharq Newspaper, January 20th, 2008

Al-Sharq Newspaper, January 20th, 2008

Al-Sharq Newspaper, January 20th, 2008

Al-Sharq Newspaper, January 21st, 2008

Al-Sharq Newspaper, January 21st, 2008

Al-Sharq Newspaper, January 22nd, 2008

Al-Sharq Newspaper, January 22nd, 2008

Al-Sharq Newspaper, January 23rd, 2008

Al-Sharq Newspaper, January 23rd, 2008

Al-Sharq Newspaper, January 24th, 2008

Al-Sharq Newspaper, January 28th, 2008

Al-Sharq Newspaper, January 29th, 2008

Al-Sharq Newspaper, January 29th, 2008

Al-Sharq Newspaper, January 29th, 2008

Al-Sharq Newspaper, January 31st, 2008

Al-Sharq Newspaper, January 31st, 2008

Al-Sharq Newspaper, January 31st, 2008

Al-Sharq Newspaper, February 1st, 2008

Al-Sharq Newspaper, February 2nd, 2008

Al-Sharq Newspaper, February 3rd, 2008

Al-Sharq Newspaper, February 3rd, 2008

Al-Sharq Newspaper, February 4th, 2008

Al-Sharq Newspaper, February 5th, 2008

Al-Sharq Newspaper, February 10th, 2008

Al-Sharq Newspaper, February 10th, 2008

Al-Sharq Newspaper, February 10th, 2008

Al-Sharq Newspaper, February 10th, 2008

Al-Sharq Newspaper, February 12th, 2008

Al-Sharq Newspaper, February 12th, 2008

Al-Sharq Newspaper, February 13th, 2008

Al-Sharq Newspaper, February 14th, 2008

Al-Sharq Newspaper, February 15th, 2008

Al-Sharq Newspaper, February 16th, 2008

Al-Sharq Newspaper, February 17th, 2008

Al-Sharq Newspaper, February 17th, 2008

Al-Sharq Newspaper, February 17th, 2008

Al-Sharq Newspaper, February 17th, 2008

Al-Sharq Newspaper, February 18th, 2008

Al-Sharq Newspaper, February 18th, 2008

Al-Sharq Newspaper, February 18th, 2008

Al-Sharq Newspaper, February 18th, 2008

Al-Sharq Newspaper, February 19th, 2008

Al-Sharq Newspaper, February 21st, 2008

Al-Sharq Newspaper, February 21st, 2008, page 7

Al-Sharq Newspaper, February 21st, 2008, page 27

Al-Sharq Newspaper, February 24th, 2008, Educational Sharq, No. 10

Al-Sharq Newspaper, February 24th, 2008, page 5

Al-Sharq Newspaper, February 24th, 2008

Al-Sharq Newspaper, February 24th, 2008

Al-Sharq Newspaper, February 25th, 2008

Al-Sharq Newspaper, February 25th, 2008

Al-Sharq Newspaper, February 26th, 2008

Al-Sharq Newspaper, February 27th, 2008

Al-Sharq Newspaper, February 27th, 2008

Al-Sharq Newspaper, February 27th, 2008, page 35

Al-Sharq Newspaper, February 27th, 2008, page 35

Al-Sharq Newspaper, February 28th, 2008

Al-Sharq Newspaper, February 28th, 2008

Al-Sharq Newspaper, February 28th, 2008

Al-Sharq Newspaper, February 28th, 2008

Al-Sharq Newspaper, March 1st, 2008

Al-Sharq Newspaper, March 2nd, 2008

Al-Sharq Newspaper, March 3rd, 2008

Al-Sharq Newspaper, March 3rd, 2008

Al-Sharq Newspaper, March 4th, 2008

Al-Sharq Newspaper, March 5th, 2008

Al-Sharq Newspaper, March 6th, 2008

Al-Sharq Newspaper, March 6th, 2008

Al-Sharq Newspaper, March 9th, 2008

Al-Sharq Newspaper, March 9th, 2008, Educational Sharq, No 12

Al-Sharq Newspaper, March 9th, 2008

Al-Sharq Newspaper, March 10th, 2008

Al-Sharq Newspaper, March 10th, 2008

Al-Sharq Newspaper, March 11th, 2008

Al-Sharq Newspaper, March 11th, 2008

Al-Sharq Newspaper, March 11th, 2008

Al-Sharq Newspaper, March 11th, 2008

Al-Sharq Newspaper, March 12th, 2008

Al-Sharq Newspaper, March 13th, 2008

Al-Sharq Newspaper, March 13th, 2008

Al-Sharq Newspaper, March 13th, 2008

Al-Sharq Newspaper, March 13th, 2008

Al-Sharq Newspaper, March 16th, 2008

Al-Sharq Newspaper, March 17th, 2008

Al-Sharq Newspaper, March 18th, 2008

Al-Sharq Newspaper, March 18th, 2008

Al-Sharq Newspaper, March 18th, 2008

Al-Sharq Newspaper, March 18th, 2008

Al-Sharq Newspaper, March 19th, 2008

Al-Sharq Newspaper, March 20th, 2008

Al-Sharq Newspaper, March 20th, 2008

Al-Sharq Newspaper, March 23rd, 2008

Al-Sharq Newspaper, March 23rd, 2008

Al-Sharq Newspaper, March 23rd, 2008

Al-Sharq Newspaper, March 24th, 2008

Al-Sharq Newspaper, March 24th, 2008

Al-Sharq Newspaper, March 25th, 2008

Al-Sharq Newspaper, March 25th, 2008

Al-Sharq Newspaper, March 26th, 2008

Al-Sharq Newspaper, March 26th, 2008

Al-Sharq Newspaper, March 27th, 2008

Al-Sharq Newspaper, March 27th, 2008

Al-Sharq Newspaper, March 30th, 2008

Al-Sharq Newspaper, March 30th, 2008

Al-Sharq Newspaper, March 30th, 2008

Al-Sharq Newspaper, April 1st, 2008

Al-Sharq Newspaper, April 1st, 2008

Al-Sharq Newspaper, April 1st, 2008

Al-Sharq Newspaper, April 1st, 2008

Al-Sharq Newspaper, April 3rd, 2008

Al-Sharq Newspaper, April 6th, 2008

Al-Sharq Newspaper, April 6th, 2008, Educational Sharq

Al-Sharq Newspaper, April 7th, 2008

Al-Sharq Newspaper, April 8th, 2008

Al-Sharq Newspaper, April 8th, 2008

Al-Sharq Newspaper, April 8th, 2008

Al-Sharq Newspaper, April 9th, 2008

Al-Sharq Newspaper, April 13th, 2008

Al-Sharq Newspaper, April 13th, 2008

Al-Sharq Newspaper, April 13th, 2008

Al-Sharq Newspaper, April 14th, 2008

Al-Sharq Newspaper, April 14th, 2008

Al-Sharq Newspaper, April 16th, 2008

Al-Sharq Newspaper, April 17th, 2008

Al-Sharq Newspaper, April 17th, 2008

Al-Sharq Newspaper, April 17th, 2008

Al-Sharq Newspaper, April 20th, 2008

Al-Sharq Newspaper, April 21st, 2008

Al-Sharq Newspaper, April 21st, 2008

Al-Sharq Newspaper, April 21st, 2008

Al-Sharq Newspaper, April 21st, 2008

Al-Sharq Newspaper, April 22nd, 2008

Al-Sharq Newspaper, April 27th, 2008

Al-Sharq Newspaper, April 28th, 2008

Al-Sharq Newspaper, April 30th, 2008

Al-Sharq Newspaper, April 30th, 2008

Al-Sharq Newspaper, May 1st, 2008

Al-Sharq Newspaper, May 1st, 2008

Al-Sharq Newspaper, May 4th, 2008

Al-Sharq Newspaper, May 4th, 2008

Al-Sharq Newspaper, May 5th, 2008

Al-Sharq Newspaper, May 5th, 2008

Al-Sharq Newspaper, May 6th, 2008

Al-Sharq Newspaper, May 7th, 2008

Al-Sharq Newspaper, May 7th, 2008

Al-Sharq Newspaper, May 8th, 2008

Al-Sharq Newspaper, May 8th, 2008

Al-Sharq Newspaper, May 11th, 2008

Al-Sharq Newspaper, May 11th, 2008

Al-Sharq Newspaper, May 11th, 2008

Al-Sharq Newspaper, May 13th, 2008

Al-Sharq Newspaper, May 13th, 2008

Al-Sharq Newspaper, May 14th, 2008

Al-Sharq Newspaper, May 15th, 2008

Al-Sharq Newspaper, May 18th, 2008

Al-Sharq Newspaper, May 18th, 2008

Al-Sharq Newspaper, May 18th, 2008

Al-Sharq Newspaper, May 19th, 2008

Al-Sharq Newspaper, May 20th, 2008

Al-Sharq Newspaper, May 21st, 2008

Al-Sharq Newspaper, May 21st, 2008

Al-Sharq Newspaper, May 21st, 2008

Al-Sharq Newspaper, May 25th, 2008

Al-Sharq Newspaper, May 25th, 2008

Al-Sharq Newspaper, May 25th, 2008

Al-Sharq Newspaper, May 25th, 2008, Educational Sharq

Al-Sharq Newspaper, May 26th, 2008

Al-Sharq Newspaper, May 27th, 2008

Al-Sharq Newspaper, May 27th, 2008

Al-Sharq Newspaper, May 27th, 2008

Al-Sharq Newspaper, May 27th, 2008

Al-Sharq Newspaper, May 29th, 2008

Al-Sharq Newspaper, May 29th, 2008

Al-Sharq Newspaper, May 29th, 2008

Al-Watan Newspaper, September 7th, 2004

Al-Watan Newspaper, September 9th, 2004

Al-Watan Newspaper, September 20th, 2004

Al-Watan Newspaper, September 21st, 2004

Al-Watan Newspaper, September 23rd, 2004

Al-Watan Newspaper, September 28th, 2004

Al-Watan Newspaper, October 12th, 2004

Al-Watan Newspaper, October 24th, 2004

Al-Watan Newspaper, November 8th, 2004

Al-Watan Newspaper, November 21st, 2004

Al-Watan Newspaper, November 21st, 2004

Al-Watan Newspaper, December 2nd, 2004

Al-Watan Newspaper, December 14th, 2004

Al-Watan Newspaper, December 21st, 2004

Al-Watan Newspaper, January 2nd, 2005

Al-Watan Newspaper, January 17th, 2005

Al-Watan Newspaper, January 17th, 2005

Al-Watan Newspaper, January 26th, 2005

Al-Watan Newspaper, February 16th, 2005

Al-Watan Newspaper, February 21st, 2005

Al-Watan Newspaper, February 27th, 2005

Al-Watan Newspaper, February 27th, 2005

Al-Watan Newspaper, February 28th, 2005

Al-Watan Newspaper, February 28th, 2005

Al-Watan Newspaper, March 8th, 2005

Al-Watan Newspaper, March 16th, 2005

Al-Watan Newspaper, March 23rd, 2005

Al-Watan Newspaper, April 4th, 2005

Al-Watan Newspaper, April 7th, 2005

Al-Watan Newspaper, April 10th, 2005

Al-Watan Newspaper, April 10th, 2005

Al-Watan Newspaper, April 10th, 2005

Al-Watan Newspaper, April 12th, 2005

Al-Watan Newspaper, April 13th, 2005

Al-Watan Newspaper, April 17th, 2005

Al-Watan Newspaper, April 19th, 2005

Al-Watan Newspaper, April 26th, 2005

Al-Watan Newspaper, April 26th, 2005

Al-Watan Newspaper, April 29th, 2005

Al-Watan Newspaper, May 1st, 2005

Al-Watan Newspaper, May 2nd, 2005

Al-Watan Newspaper, May 19th, 2005

Al-Watan Newspaper, May 22nd, 2005

Al-Watan Newspaper, May 31st, 2005

Al-Watan Newspaper, June 1st, 2005

Al-Watan Newspaper, June 20th, 2005

Al-Watan Newspaper, June 27th, 2005

Al-Watan Newspaper, June 3rd, 2007

Al-Watan Newspaper, June 4th, 2007

Al-Watan Newspaper, June 5th, 2007

Al-Watan Newspaper, June 5th, 2007

Al-Watan Newspaper, June 6th, 2007

Al-Watan Newspaper, June 6th, 2007

Al-Watan Newspaper, June 6th, 2007

Al-Watan Newspaper, June 7th, 2007

Al-Watan Newspaper, June 10th, 2007

Al-Watan Newspaper, June 10th, 2007

Al-Watan Newspaper, June 10th, 2007

Al-Watan Newspaper, June 10th, 2007

Al-Watan Newspaper, June 11th, 2007

Al-Watan Newspaper, June 12th, 2007

Al-Watan Newspaper, June 12th, 2007

Al-Watan Newspaper, June 13th, 2007

Al-Watan Newspaper, June 13th, 2007

Al-Watan Newspaper, June 17th, 2007

Al-Watan Newspaper, June 19th, 2007

Al-Watan Newspaper, June 20th, 2007

Al-Watan Newspaper, June 21st, 2007

Al-Watan Newspaper, June 24th, 2007

Al-Watan Newspaper, June 24th, 2007

Al-Watan Newspaper, June 24th, 2007

Al-Watan Newspaper, June 24th, 2007

Al-Watan Newspaper, June 25th, 2007

Al-Watan Newspaper, June 25th, 2007

Al-Watan Newspaper, June 26th, 2007

Al-Watan Newspaper, June 26th, 2007

Al-Watan Newspaper, June 27th, 2007

Al-Watan Newspaper, June 28th, 2007

Al-Watan Newspaper, August 8th, 2007

Al-Watan Newspaper, August 8th, 2007

Al-Watan Newspaper, August 13th, 2007

Al-Watan Newspaper, August 22nd, 2007

Al-Watan Newspaper, August 23rd, 2007

Al-Watan Newspaper, August 30th, 2007

Al-Watan Newspaper, September 4th, 2007

Al-Watan Newspaper, September 5th, 2007

Al-Watan Newspaper, September 5th, 2007

Al-Watan Newspaper, September 6th, 2007

Al-Watan Newspaper, September 6th, 2007

Al-Watan Newspaper, September 9th, 2007

Al-Watan Newspaper, September 10th, 2007

Al-Watan Newspaper, September 11th, 2007

Al-Watan Newspaper, September 12th, 2007

Al-Watan Newspaper, September 12th, 2007

Al-Watan Newspaper, September 12th, 2007

Al-Watan Newspaper, September 20th, 2007

Al-Watan Newspaper, September 26th, 2007

Al-Watan Newspaper, September 30th, 2007

Al-Watan Newspaper, September 30th, 2007

Al-Watan Newspaper, October 2nd, 2007

Al-Watan Newspaper, October 2nd, 2007

Al-Watan Newspaper, October 3rd, 2007

Al-Watan Newspaper, October 4th, 2007

Al-Watan Newspaper, October 4th, 2007

Al-Watan Newspaper, October 4th, 2007

Al-Watan Newspaper, October 7th, 2007

Al-Watan Newspaper, October 7th, 2007

Al-Watan Newspaper, October 9th, 2007

Al-Watan Newspaper, October 23rd, 2007

Al-Watan Newspaper, October 24th, 2007

Al-Watan Newspaper, October 25th, 2007

Al-Watan Newspaper, October 28th, 2007

Al-Watan Newspaper, October 28th, 2007

Al-Watan Newspaper, October 29th, 2007

Al-Watan Newspaper, October 30th, 2007

Al-Watan Newspaper, October 31st, 2007

Al-Watan Newspaper, November 1st, 2007

Al-Watan Newspaper, November 4th, 2007

Al-Watan Newspaper, November 5th, 2007

Al-Watan Newspaper, November 6th, 2007

Al-Watan Newspaper, November 8th, 2007

Al-Watan Newspaper, November 8th, 2007

Al-Watan Newspaper, November 11th, 2007

Al-Watan Newspaper, November 11th, 2007

Al-Watan Newspaper, November 11th, 2007

Al-Watan Newspaper, November 11th, 2007

Al-Watan Newspaper, November 12th, 2007

Al-Watan Newspaper, November 12th, 2007

Al-Watan Newspaper, November 15th, 2007

Al-Watan Newspaper, November 18th, 2007

Al-Watan Newspaper, November 18th, 2007

Al-Watan Newspaper, November 20th, 2007

Al-Watan Newspaper, November 20th, 2007

Al-Watan Newspaper, November 21st, 2007

Al-Watan Newspaper, November 27th, 2007

Al-Watan Newspaper, November 29th, 2007

Al-Watan Newspaper, December 3rd, 2007

Al-Watan Newspaper, December 4th, 2007

Al-Watan Newspaper, December 4th, 2007

Al-Watan Newspaper, December 6th, 2007

Al-Watan Newspaper, December 10th, 2007

Al-Watan Newspaper, December 11th, 2007

Al-Watan Newspaper, December 13th, 2007

Al-Watan Newspaper, December 16th, 2007

Al-Watan Newspaper, December 17th, 2007

Al-Watan Newspaper, December 17th, 2007

Al-Watan Newspaper, December 26th, 2007

Al-Watan Newspaper, December 27th, 2007

Al-Watan Newspaper, December 30th, 2007

Al-Watan Newspaper, December 31st, 2007

Al-Watan Newspaper, December 31st, 2007

Al-Watan Newspaper, January 1st, 2008

Al-Watan Newspaper, January 6th, 2008

Al-Watan Newspaper, January 7th, 2008

Al-Watan Newspaper, January 7th, 2008

Al-Watan Newspaper, January 10th, 2008

Al-Watan Newspaper, January 10th, 2008

Al-Watan Newspaper, January 13th, 2008

Al-Watan Newspaper, January 13th, 2008

Al-Watan Newspaper, January 14th, 2008

Al-Watan Newspaper, January 16th, 2008

Al-Watan Newspaper, January 16th, 2008

Al-Watan Newspaper, January 17th, 2008

Al-Watan Newspaper, January 17th, 2008

Al-Watan Newspaper, January 20th, 2008

Al-Watan Newspaper, January 21st, 2008

Al-Watan Newspaper, January 22nd, 2008

Al-Watan Newspaper, January 22nd, 2008

Al-Watan Newspaper, January 22nd, 2008

Al-Watan Newspaper, January 23rd, 2008

Al-Watan Newspaper, January 28th, 2008

Al-Watan Newspaper, January 29th, 2008

Al-Watan Newspaper, January 31st, 2008

Al-Watan Newspaper, January 31st, 2008

Al-Watan Newspaper, February 3rd, 2008

Al-Watan Newspaper, February 13th, 2008

Al-Watan Newspaper, February 13th, 2008

Al-Watan Newspaper, February 13th, 2008

Al-Watan Newspaper, February 17th, 2008

Al-Watan Newspaper, February 19th, 2008

Al-Watan Newspaper, February 19th, 2008

Al-Watan Newspaper, February 19th, 2008

Al-Watan Newspaper, February 21st, 2008

Al-Watan Newspaper, February 24th, 2008

Al-Watan Newspaper, February 26th, 2008

Al-Watan Newspaper, February 26th, 2008

Al-Watan Newspaper, February 27th, 2008

Al-Watan Newspaper, February 27th, 2008

Al-Watan Newspaper, February 28th, 2008

Al-Watan Newspaper, February 28th, 2008

Al-Watan Newspaper, February 28th, 2008, page 8

Al-Watan Newspaper, February 28th, 2008, page 12

Al-Watan Newspaper, March 2nd, 2008

Al-Watan Newspaper, March 4th, 2008

Al-Watan Newspaper, March 4th, 2008

Al-Watan Newspaper, March 6th, 2008

Al-Watan Newspaper, March 6th, 2008

Al-Watan Newspaper, March 9th, 2008

Al-Watan Newspaper, March 9th, 2008

Al-Watan Newspaper, March 10th, 2008

Al-Watan Newspaper, March 10th, 2008

Al-Watan Newspaper, March 10th, 2008

Al-Watan Newspaper, March 11th, 2008

Al-Watan Newspaper, March 11th, 2008

Al-Watan Newspaper, March 11th, 2008

Al-Watan Newspaper, March 13th, 2008

Al-Watan Newspaper, March 13th, 2008

Al-Watan Newspaper, March 17th, 2008

Al-Watan Newspaper, March 18th, 2008

Al-Watan Newspaper, March 18th, 2008

Al-Watan Newspaper, March 19th, 2008

Al-Watan Newspaper, March 20th, 2008

Al-Watan Newspaper, March 20th, 2008

Al-Watan Newspaper, March 20th, 2008

Al-Watan Newspaper, March 24th, 2008

Al-Watan Newspaper, March 24th, 2008

Al-Watan Newspaper, March 24th, 2008

Al-Watan Newspaper, March 25th, 2008

Al-Watan Newspaper, March 25th, 2008

Al-Watan Newspaper, March 25th, 2008

Al-Watan Newspaper, March 26th, 2008

Al-Watan Newspaper, March 27th, 2008

Al-Watan Newspaper, March 27th, 2008

Al-Watan Newspaper, March 30th, 2008

Al-Watan Newspaper, April 1st, 2008

Al-Watan Newspaper, April 1st, 2008

Al-Watan Newspaper, April 1st, 2008

Al-Watan Newspaper, April 1st, 2008

Al-Watan Newspaper, April 1st, 2008

Al-Watan Newspaper, April 3rd, 2008

Al-Watan Newspaper, April 6th, 2008

Al-Watan Newspaper, April 13th, 2008

Al-Watan Newspaper, April 14th, 2008

Al-Watan Newspaper, April 14th, 2008

Al-Watan Newspaper, April 15th, 2008

Al-Watan Newspaper, April 15th, 2008

Al-Watan Newspaper, April 16th, 2008

Al-Watan Newspaper, April 16th, 2008

Al-Watan Newspaper, April 17th, 2008

Al-Watan Newspaper, April 20th, 2008

Al-Watan Newspaper, April 21st, 2008

Al-Watan Newspaper, April 21st, 2008

Al-Watan Newspaper, April 21st, 2008

Al-Watan Newspaper, April 22nd, 2008

Al-Watan Newspaper, April 27th, 2008

Al-Watan Newspaper, April 28th, 2008

Al-Watan Newspaper, April 28th, 2008

Al-Watan Newspaper, April 30th, 2008

Al-Watan Newspaper, April 30th, 2008

Al-Watan Newspaper, April 30th, 2008

Al-Watan Newspaper, May 1st, 2008

Al-Watan Newspaper, May 1st, 2008

Al-Watan Newspaper, May 1st, 2008

Al-Watan Newspaper, May 1st, 2008

Al-Watan Newspaper, May 5th, 2008

Al-Watan Newspaper, May 5th, 2008

Al-Watan Newspaper, May 6th, 2008

Al-Watan Newspaper, May 7th, 2008

Al-Watan Newspaper, May 8th, 2008

Al-Watan Newspaper, May 8th, 2008

Al-Watan Newspaper, May 8th, 2008

Al-Watan Newspaper, May 13th, 2008

Al-Watan Newspaper, May 14th, 2008

Al-Watan Newspaper, May 14th, 2008

Al-Watan Newspaper, May 20th, 2008

Al-Watan Newspaper, May 21st, 2008

Al-Watan Newspaper, May 21st, 2008

Al-Watan Newspaper, May 21st, 2008

Al-Watan Newspaper, May 25th, 2008

Al-Watan Newspaper, May 25th, 2008

Al-Watan Newspaper, May 26th, 2008

Al-Watan Newspaper, May 26th, 2008

Al-Watan Newspaper, May 27th, 2008

Al-Watan Newspaper, May 29th, 2008

Al-Watan Newspaper, May 29th, 2008

Brewer, Dominic J., Catherine H. Augustine, Gail L. Zelman, Gery Ryan, Charles A. Goldman, Cathleen Stasz, and Louay Constant, Education for a new Era: Design and Implementation of K-12 Education Reform in Qatar, MG-548-Qatar, Santa Monica, Ca: RAND Corporation, 2007

Gonzalez, Gabriella, Vi-Nhuan le, Markus Broer, Louis T. Mariano, J. Enrique Froemel, Charles Goldman, and Julie DaVanzo, Lessons from the Field: Developing and Implementing the Qatar Student Assessment System, 2002-2006, TR-620-QATAR, Santa Monica, Ca: RAND Corporation, 2009

Supreme Education Council, Schools and Schooling in Qatar: 2008-09 Report, a statistical overview of aspects of schools and schooling in Qatar. Doha December 2009

Supreme Education Council, Education Reform Becomes Reality: Building the Future of Qatar, 2004 Annual Report, Doha, 2005

Supreme Education Council, 2005 Annual Report, Doha 2006

Supreme Education Council, 2006 Annual Report, Doha 2007

Supreme Education Council, 2007 Annual Report, Doha 2008

Supreme Education Council, 2008 Annual Report, Doha 2009

Supreme Education Council, Education for a New Era: From Vision to Reality 2002-2007, Report by the Office of Communication, 2008

Supreme Education Council, "SEC Official Website," As of December, 2006:
http://www.english.education.gov.qa/section/sec/

Supreme Education Council, "Qatar Comprehensive Educational Assessment (QCEA)," As of February, 2007:
http://www.english.education.gov.qa/section/sec/evaluation_institute/assessment/_qcea2005

Supreme Education Council, "Qatar Comprehensive Educational Assessment (QCEA)," As of October, 2007:
http://www.english.education.gov.qa/section/sec/evaluation_institute/assessment/_qcea_2006

Supreme Education Council, "Qatar Comprehensive Educational Assessment (QCEA)," As of November, 2008:
http://www.english.education.gov.qa/section/sec/evaluation_institute/assessment/_qcea_2007

Supreme Education Council, "School Report Cards (SRC)," As of December, 2006:
http://www.education.gov.qa/SRC/search0405.htm

Supreme Education Council, "School Report Cards (SRC)," As of October, 2007:
http://www.english.education.gov.qa/SRC/0506/src0506.htm

Supreme Education Council, "School Report Cards (SRC)," As of November, 2008:
http://www.english.education.gov.qa/SRC/0607/src0607.htm

Supreme Education Council, "Review of Schools," As of August, 2007:
http://www.english.education.gov.qa/section/sec/evaluation_institute/seo/_is_review

Watani Al-Habeeb Radio Program broadcast date: November 30th, 2004
Watani Al-Habeeb Radio Program broadcast date: December 14th, 2004
Watani Al-Habeeb Radio Program broadcast date: December 15th 2004
Watani Al-Habeeb Radio Program broadcast date: December 23rd 2004
Watani Al-Habeeb Radio Program broadcast date: January 1st 2005
Watani Al-Habeeb Radio Program broadcast date: January 3rd 2005
Watani Al-Habeeb Radio Program broadcast date: January 3rd 2005
Watani Al-Habeeb Radio Program broadcast date: January 12th 2005
Watani Al-Habeeb Radio Program broadcast date: January 24th 2005
Watani Al-Habeeb Radio Program broadcast date: April 6th 2005
Watani Al-Habeeb Radio Program broadcast date: April 10th, 2005
Watani Al-Habeeb Radio Program broadcast date: April 10th, 2005
Watani Al-Habeeb Radio Program broadcast date: April 10th, 2005

Watani Al-Habeeb Radio Program broadcast date: April 13th, 2005

Watani Al-Habeeb Radio Program broadcast date: April 20th, 2005

Watani Al-Habeeb Radio Program broadcast date: April 25th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 4th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 10th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 10th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 16th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 17th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 18th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 18th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 18th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 18th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 18th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 22nd, 2005

Watani Al-Habeeb Radio Program broadcast date: May 23rd, 2005

Watani Al-Habeeb Radio Program broadcast date: May 25th, 2005

Watani Al-Habeeb Radio Program broadcast date: May 25th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 6th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 6th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 7th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 8th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 13th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 13th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 14th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 15th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 15th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 15th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 16th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 16th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 19th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 22nd, 2005

Watani Al-Habeeb Radio Program broadcast date: June 22nd, 2005

Watani Al-Habeeb Radio Program broadcast date: June 23rd, 2005

Watani Al-Habeeb Radio Program broadcast date: June 26th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 27th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 27th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 28th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 28th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 30th, 2005

Watani Al-Habeeb Radio Program broadcast date: June 4th, 2007

Watani Al-Habeeb Radio Program broadcast date: June 6th, 2007

Watani Al-Habeeb Radio Program broadcast date: June 12th, 2007

Watani Al-Habeeb Radio Program broadcast date: June 18th, 2007

Watani Al-Habeeb Radio Program broadcast date: March 26th, 2008

Watani Al-Habeeb Radio Program broadcast date: March 30th, 2008

APPENDIX 2

INTERVIEW PROTOCOL AND QUESTIONS

INTERVIEW PROTOCOL

1. A standard brief description of the research, name of the researcher, academic affiliation, and the purpose of the meeting to be drafted and communicated to all interviewees or their assistants when the researcher calls in to take an appointment for the interview;

2. A cover letter that lists date, the interviewee's first and last name initials, and place of interview is printed out for each interview. The cover letter is stapled together with a copy of the interview questions form and some blank sheets of paper;

3. Each interview commences with reading the oral consent form and addressing any concerns/questions that the interviewee might raise before proceeding to asking the questions;

4. The interviewer starts to ask the questions in the order that they are listed on the list. Notes are written in the space provided underneath each question;

5. The interviewer can use a notepad if the space provided is not enough to record the response. Every effort should be made to write the number of the question on the notepad prior to recording the response;

6. Where applicable, and as noted on the form, in the order listed, to provide reference to the main themes, ensure that the interviewee understand the points being made, gather specific examples and get further insights;

7. For open-ended questions that are followed by another set of questions, the researcher will first listen to the interviewee's response before starting to ask the sub-questions in the order that they are listed;

8. After conducting each interview, notes are to be written up within the next 24 hours;

9. The researcher to note down any missing information, unclear answers, and other points that need further clarifications or data collection;

10. All interviewees who participated in the personal interviews are to be asked to validate their responses. A draft of the interview transcripts is to be sent to each interviewee filled with his/her own responses;

11. Another appointment is scheduled with each of these interviewees after sending them a copy of the transcripts in order to discuss their comments and feedback as well as ask additional questions if necessary;

12. Once responses are validated and additional questions are asked, revised interview transcripts are produced for the personal interviews and immediately used as the basis for analysis.

STAKEHOLDERS INTERVIEW QUESTIONS

[Introductions and Oral Consent Protocol]

A. THE PERFORMANCE MEASURES

1. Do you receive any data from the National Education Data System? What exactly do you receive?

- (If the answer is no, thank the interviewee and end the interview)
- Probe: results of the educational appraisal scheme, results of the school appraisal scheme, school articles, copies of what is published in the media

2. Do you know the other recipients of this National Education System data?

- Probe: within your organization/ outside your organization. Locally/regionally/internationally.

3. Are there any entities/organizations/people that should be receiving this data but are currently not?

- Probe: these could be parties that have interest in the reform and its progress/ are affected, directly or indirectly by the reform/ have power to support or change the track of the reform.

4. How often do you receive data from the National Education Data System?

- Probe: is there a timetable that is followed all the time or is data received on random basis.

5. How is the data delivered to you? In what form is the data delivered?

- Probe for the form: general raw data in open files, specific and relevant data in open files, data with some analysis/comparison, data with tables/graphic illustrations, data is delivered as part of a report, data is delivered as part of a standardized report.

6. Do you receive data about the same issues/topics every time or do covered issues/topics change from time to time? If they change who or what causes that change? Is the change ever initiated by you?

7. What do you think about the quality of the data you receive?

- Clarity of data
- Organization of the data
- Relevance of data to receiver
- Brevity and focus
- Level of details
- Level of analysis
- Comparison with past data
- Quality of illustration

8. What do you think about the National Education Data System in general?

- Probe: not familiar with its characteristics, satisfied: the system meets your needs, unsatisfied: the system needs improvement.
- Probe: any comments about data collection methods, validity of data, data analysis techniques.

9. Do you have any ideas or proposals to improve the National Education data system?

- Probe for specific ideas:
    - Tailor to specific needs of users
    - Get periodic feedback from users
    - Covers too many areas; spread thin over irrelevant issues
    - Link to priorities and specific goals
    - Develop staff capacity/Improve scientific methods used to collect and analyze data
    - Improve data reporting
    - Modernize IT system

B. THE UTILIZATION PROCESS

10. Do you use the data you receive from the national education data system? If yes, do you use it on regular basis? How often?

11. What do you use the data you receive for? Please give examples.

- Probe for specific uses:
    - to monitor performance
    - to evaluate progress
    - to better plan for the future
    - to prepare reports
    - to guide decision-making at an operational level (staffing, buildings, curriculum, student/ratio ratios)
    - to guide decision-making at a strategic level (strategic planning)
    - to affect policy changes
    - to prepare for senior management/board meetings
    - to prepare for presentations/public briefings.

12. How do you use the data? Is there a specific way/procedure/system that is followed/adhered to during the process of using the data? Please describe that way/process?

Probe: There is no system: relevant data is pulled out and used when and if needed. There is a documented & institutionalized step-by-step procedure that employees need to follow when using the data. There are employees who are specifically charged with reviewing/reporting on the data, using it for our specific needs, following up on it specific use.

13. How do you share knowledge derived from the data within your immediate entity and with other entities?

Probe: emails, informal meeting, formal meetings, reports, briefings, others.

C. FACTORS AFFECTING UTILIZATION

14. Do you think you have the right set of enablers that allow you to fully utilize the data you receive and take advantage of it?

Probe: sufficient financial and human resources, technical knowledge, access to the right kind of data, clear goals and consensus on strategy and direction.

15. Do you think you that the overall cultural and political context allows you to fully utilize the data you receive and take advantage of it?

Probe: culture and attitude, internal interest groups, external interest groups, the media

16. Are there any other factors that are facilitating or inhibiting the process of utilization? What are these factors?

17. What do you think is the most effective way to utilize the data you receive and fully take advantage of it? What would be an ideal situation/process?

18. Do you have any other comments that you would like to add?