

**KATHOLIEKE UNIVERSITEIT LEUVEN**  
**FACULTEIT SOCIALE WETENSCHAPPEN**

# **Performance Measurement in the Flemish Public Sector: A Supply and Demand Approach**

**Wouter VAN DOOREN**

Proefschrift tot het verkrijgen van de graad van  
Doctor in de Sociale Wetenschappen

Nr. 96

Samenstelling van de examencommissie:

prof. dr. G. Bouckaert (promotor), prof. dr. M. Aristigueta (University of Delaware, U.S.A.),  
prof. dr. M. Brans, prof. dr. P. Laegreid (University of Bergen, Norway), prof. dr. W. Moesen;  
prof. dr. K. Roe (voorzitter), dr. J. De la Haye (secretaris)

2006

De verantwoordelijkheid voor de ingenomen standpunten berust alleen bij de auteur.

Gepubliceerd door:

Faculteit Sociale Wetenschappen - Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven,  
E. Van Evenstraat 2A - 3000 Leuven, België.

© 2006 by the author.

Niets uit deze uitgave mag worden veeelvoudigd zonder voorafgaande schriftelijke toestemming van de  
auteur / No part of this book may be reproduced in any form without the permission in writing from the  
author.

D/2006/8978/2

## Preface

*"We seem to be able to land on Mars, but not on Earth."* John Logsdon, director of the Space Policy Institute, made a very accurate analysis when he referred to the crash landing of NASA's Genesis space probe in the Utah desert (8/9/2004). Yet, some clemency towards NASA is appropriate. To land on both Mars and back on Earth is an immense undertaking. In Public Administration too, theoretical discussions and practical applications often seem decoupled. This thesis is less ambitious. The main goal was to put some significant steps forward towards a better understanding of performance measurement and management in public administration. Therefore, middle range theorizing is combined with a solid empirical foundation. There was no aspiration to reach Mars, but we did want to get back on Earth.

Een doctoraat schrijf je niet alleen. Vooreerst wil ik Geert Bouckaert, promotor van deze thesis, willen bedanken. Zijn onderzoekservaring, in het bijzonder rond prestatiemeting, vormde het fundament waar ik op verder kon bouwen. Naast de intellectuele bijdrage, zou ik Geert ook willen bedanken voor zijn inzet om het Instituut voor de Overheid uit te bouwen tot een echt kenniscentrum voor de overheid en een professionele onderzoeksomgeving. Om te doctoreren in onze onderzoeksafdeling moet/mag je geen einzelgänger zijn. Hierbij wil ook mijn collega's bedanken voor hun steun en voor de levendige discussies, soms over bestuurskunde en soms ook niet. In het bijzonder wil ik Miekatrien Sterck bedanken, co-bureaugenoot, co-onderzoeker in het spoor financiën en prestatiemeting, en vooral mijn officieuze co-promotor. Ik wil ook mijn familie en vrienden bedanken. Ik hoop dat ze niet al te veel last gehad hebben van mijn doctoraatsplannen. Tenslotte wil ik Birgit bedanken om af en toe ook nog te denken aan wat echt telt.

## Table of contents

# PART ONE: A PUBLIC ADMINISTRATION PERSPECTIVE ON PERFORMANCE MEASUREMENT

<b>1.THE CONTEXT OF PERFORMANCE MEASUREMENT</b>	<b>13</b>
1.1. PERFORMANCE IN THE PUBLIC SECTOR	13
1.2. THE PROMISE OF PERFORMANCE IN THE PUBLIC SECTOR	16
1.3. MEASUREMENT OF PERFORMANCE IS A KEY FACET OF THE CONTEMPORARY PUBLIC SECTOR	17
1.3.1. <i>Performance information is pivotal in public sector reform</i>	17
1.3.2. <i>Performance information plays a central role in public management.</i>	19
1.3.3. <i>Performance information plays a central role in public policy cycle.</i>	20
1.4. PERFORMANCE MEASUREMENT RESEARCH IN FLANDERS	22
1.5. RESEARCH QUESTIONS	23
1.6. STRUCTURE OF THE TEXT	26
<b>2. GENERAL FRAMEWORK: WHAT IS PERFORMANCE MEASUREMENT?</b>	<b>27</b>
2.1. THE VOCABULARY OF PERFORMANCE MEASUREMENT	28
2.2. THE OBJECT OF STUDY: SUPPLY AND DEMAND OF PERFORMANCE INFORMATION	31
2.2.1. <i>Supply and demand, causes and effects</i>	31
2.2.2. <i>The production process: supply of performance information</i>	35
2.2.3. <i>Uses of performance information</i>	36
<b>3.A MEASUREMENT HISTORY OF GOVERNMENT</b>	<b>41</b>
3.1. A HISTORICAL ARGUMENT FOR STUDYING PERFORMANCE MEASUREMENT	41
3.2. RESEARCH QUESTIONS FOR THE HISTORICAL STUDY	43
3.3. OVERVIEW	44
3.4. POLITICAL ARITHMETIC	47
3.5. GERMAN UNIVERSITY STATISTICS	49
3.6. PUBLIC HEALTH MOVEMENT (VITAL STATISTICS)	51
3.7. MORAL STATISTICS	53
3.8. GERMAN HISTORICAL SCHOOL	54
3.9. CENSUS	56
3.10. SOCIAL SURVEY MOVEMENT	59
3.11. SCIENTIFIC MANAGEMENT AND THE SCIENCE OF ADMINISTRATION	61
3.12. COST ACCOUNTING	63
3.13. NEW YORK BUREAU FOR MUNICIPAL RESEARCH	65
3.14. PERFORMANCE BUDGETING, PPBS AND THEIR SUCCESSORS (MBO,ZBB, GPRA)	67
3.15. SOCIAL INDICATORS AND QUALITY OF LIFE	70

3.16.	THE QUALITY GURUS	73
3.17.	NEW PUBLIC MANAGEMENT (NPM)	75
3.18.	EVIDENCE BASED POLICIES	78
3.19.	CONCLUSIONS	79
<b>4.</b>	<b>LITERATURE STUDY</b>	<b>85</b>
4.1.	PURPOSE OF THE LITERATURE STUDY	85
4.2.	RESEARCH METHOD AND DATA FOR THE LITERATURE STUDY	85
4.3.	WHAT ARE THE MAIN RESEARCH QUESTIONS IN PERFORMANCE MEASUREMENT RESEARCH	86
4.3.1.	<i>About the research question</i>	86
4.3.2.	<i>Findings</i>	87
Set 1:	causes and conditions for performance measurement	88
Set 2:	relating design (supply) and use (demand)	89
Set 3:	effects and consequences	92
4.4.	WHAT IS THE ORIENTATION AND THE RANGE OF THE THEORIES USED FOR STUDYING PERFORMANCE MEASUREMENT?	93
4.4.1.	<i>About the research question</i>	93
4.4.2.	<i>Findings</i>	97
4.5.	WHAT ARE THE METHODS USED FOR STUDYING PERFORMANCE MEASUREMENT?	98
4.5.1.	<i>About the research question</i>	98
4.5.2.	<i>Findings</i>	98
4.6.	IS THE FOCUS ON THE CITIZEN, POLITICS, OR THE ADMINISTRATION AND DOES IT CHANGE OVER TIME?	101
4.6.1.	<i>About the research question</i>	101
4.6.2.	<i>Findings</i>	102
4.7.	DISCUSSION AND CONCLUSIONS	102
 <b>PART TWO: EMPIRICAL RESEARCH</b>		
<b>5.</b>	<b>METHODOLOGY</b>	<b>109</b>
5.1.	UNSTRUCTURED INTERVIEWS	109
5.1.1.	<i>Definition</i>	109
5.1.2.	<i>Purpose</i>	110
5.1.3.	<i>Case selection and instruments in this study</i>	111
5.2.	CONTENT ANALYSIS	112
5.2.1.	<i>Definition</i>	112
5.2.2.	<i>Purpose</i>	112
5.2.3.	<i>Case selection and instruments in this study</i>	113
5.3.	SURVEY (QUESTIONNAIRE)	113
5.3.1.	<i>Definition</i>	113
5.3.2.	<i>Purpose</i>	113

5.3.3.	<i>Case selection and instruments in this study</i>	114
5.4.	SEMI-STRUCTURED INTERVIEWS	115
5.4.1.	<i>Definition</i>	115
5.4.2.	<i>Purpose</i>	116
5.4.3.	<i>Case selection and instruments in this study</i>	116
5.5.	RELATION OF DATA SOURCES AND RESEARCH ISSUES IN THIS STUDY	116
<b>6.</b>	<b>WHAT MAKES ORGANIZATIONS MEASURE?</b>	<b>119</b>
6.1.	PROBLEM DEFINITION AND PURPOSE	120
6.2.	THE ROLE OF INFORMATION IN ORGANIZATIONS	120
6.3.	METHODOLOGY	121
6.4.	DEPENDENT VARIABLES	123
6.4.1.	<i>Adoption of performance measurement</i>	123
6.4.2.	<i>Implementation of performance measurement</i>	125
6.4.3.	<i>Calculating adoption and implementation</i>	126
6.4.4.	<i>Relating adoption and implementation</i>	128
6.5.	INDEPENDENT VARIABLES, HYPOTHESES, AND RESULTS	129
6.5.1.	<i>The measurability hypothesis</i>	129
6.5.2.	<i>The political interest hypothesis</i>	131
6.5.3.	<i>The scale hypothesis</i>	133
6.5.4.	<i>The street level discretion hypothesis</i>	134
6.5.5.	<i>The means hypothesis</i>	135
6.5.6.	<i>The goal orientation - hypothesis</i>	137
6.6.	CONCLUSIONS	138
<b>7.</b>	<b>DOES ADMINISTRATIVE SUPPLY MEET POLITICAL DEMAND? A CROSS-SECTORAL COMPARISON.</b>	<b>141</b>
7.1.	INTRODUCTION	142
7.2.	RESEARCH QUESTIONS	142
7.3.	POLICY SECTOR AS A UNIT OF ANALYSIS AND A MESO LEVEL REALITY	143
7.4.	DATA AND METHODOLOGY	144
7.5.	RESULTS	148
7.6.	DISCUSSION	154
7.7.	CONCLUSION	156
<b>8.</b>	<b>WHAT ARE THE SYSTEM REQUIREMENTS FOR DIFFERENT USES OF PERFORMANCE INFORMATION?</b>	<b>159</b>
8.1.	INTRODUCTION	160
8.2.	METHODOLOGY	160
8.3.	USES OF PERFORMANCE INFORMATION	161
8.4.	THE MEASUREMENT PROCESS WITH A DIFFERENTIATION FOR USE	169

8.4.1.	<i>Step 1 establishing the implementation mode of measurement</i>	169
8.4.2.	<i>Differentiation for purposes of performance information</i>	171
8.4.3.	<i>Step 2: definition of the measurement object</i>	171
8.4.4.	<i>Differentiation for purposes of performance information</i>	173
8.4.5.	<i>Step 3: selection of the indicators</i>	174
8.4.6.	<i>Differentiation for purposes of performance information</i>	174
8.4.7.	<i>step 4: data collection</i>	175
8.4.8.	<i>Differentiation for purposes of performance information</i>	176
8.4.9.	<i>Step 5 analysis (from performance data to performance information)</i>	177
8.4.10.	<i>Differentiation for purposes of performance information</i>	181
8.4.11.	<i>Step 6: reporting</i>	182
8.4.12.	<i>Differentiation for purposes of performance information</i>	183
8.4.13.	<i>Throughout the process: quality assurance</i>	184
8.4.14.	<i>Differentiation for purposes of performance information</i>	185
8.5.	CONCLUSION	187
<b>9.WHAT ARE THE EFFECTS OF PERFORMANCE MEASUREMENT?</b>		<b>191</b>
9.1.	INTRODUCTION	192
9.2.	A MULTITUDE OF EFFECTS	193
9.2.1.	<i>Manipulation of measurement</i>	193
9.2.2.	<i>Manipulation of output</i>	195
9.3.	GOAL DISPLACEMENT; FOR GOOD AND FOR BAD	198
9.4.	CENTRAL HYPOTHESIS: USE DETERMINES EFFECTS.	200
9.5.	DATA	203
9.6.	RESULTS	208
9.6.1.	<i>triangulation of the user profiles</i>	208
9.6.2.	<i>Which effects are showing?</i>	212
9.6.3.	<i>Linking user profiles and effects</i>	214
9.7.	OPPORTUNITY AND MOTIVE	218
9.8.	CONCLUSION	222
<b>PART 3: CONCLUSIONS</b>		
<b>10.CONCLUSIONS</b>		<b>227</b>
10.1.	CONCLUSIONS FROM THE HISTORICAL STUDY OF MEASUREMENT IN GOVERNMENT	227
10.2.	CONCLUSIONS FROM THE LITERATURE STUDY	228
10.3.	CONCLUSIONS FROM THE FIRST EMPIRICAL RESEARCH CHAPTER: WHAT MAKES ORGANIZATIONS MEASURE?	230
10.4.	CONCLUSIONS FROM THE SECOND RESEARCH CHAPTER: DOES ADMINISTRATIVE SUPPLY MEET POLITICAL DEMAND?	231

10.5.	CONCLUSIONS FROM THE THIRD RESEARCH CHAPTER: WHAT ARE THE SYSTEM REQUIREMENTS FOR DIFFERENT USES OF PERFORMANCE INFORMATION?	232
10.6.	CONCLUSIONS FROM THE FOURTH RESEARCH CHAPTER: WHAT ARE THE EFFECTS OF PERFORMANCE MEASUREMENT?	232
<b>11.</b>	<b>PROSPECTIVE DIRECTIONS</b>	<b>235</b>
11.1.	A COMBINATION OF THEORIES OF INFORMATION AND THEORIES ON ORGANIZATION	235
11.1.1.	<i>Extension of the view of organizations.</i>	236
11.1.2.	<i>Extensions in the view of information: Knowledge Utilization.</i>	238
11.2.	THE POLICY SECTOR AS AN INTERMEDIATE LEVEL OF ANALYSIS	242
11.3.	ACCOUNTING FOR MACRO LEVEL FACTORS.	244
<b>12.</b>	<b>REFERENCES</b>	<b>247</b>
<b>13.</b>	<b>APPENDICES</b>	<b>281</b>
13.1.	APPENDICES TO THE LITERATURE STUDY	281
13.1.1.	<i>List of articles</i>	281
13.1.2.	<i>Analysis of the studies: research questions and conclusions</i>	286
13.1.3.	<i>Classification of the studies: focus, theoretical orientation, theoretical range and method</i>	316
13.2.	DESCRIPTIVE STATISTICS FOR CHAPTER 6	319
13.3.	DESCRIPTIVE STATISTICS FOR CHAPTER 7	327
13.4.	QUESTIONNAIRE	331
<b>SAMENVATTING</b>		<b>350</b>
<b>RESUME</b>		<b>351</b>



## Figures

Figure 1: measurement in organizational management: indications from the Belgian public sector	20
Figure 2: number of indicators in the policy plans of the ministers of the Flemish government	22
Figure 3: the input output model	28
Figure 4: research questions regarding performance measurement	34
Figure 5: the production process of performance information (supply)	35
Figure 6: a chronology of movements	46
Figure 7: the NPM triangle: reforms, societal context and intellectual frameworks	76
Figure 8: a categorization of variables in performance measurement research	87
Figure 9: set 1: causes and conditions for performance measurement	88
Figure 10: set 2: supply and demand	89
Figure 11: set 3: consequences and effects of performance measurement	92
Figure 12: small, middle and broad range theories.	95
Figure 13: characteristics of Middle Range Theories	96
Figure 14: range of the performance measurement studies	97
Figure 15: methodologies of performance measurement articles	99
Figure 16: obtrusiveness and No of observations of performance measurement studies	101
Figure 17: organizational chart of the Ministry of the Flemish Community	111
Figure 18: development process of the questionnaire	114
Figure 19: relation of data sources and research questions	117
Figure 20: outline of the study	121
Figure 21: organizational chart of the Ministry of the Flemish Community, Belgium	121
Figure 22: adoption and implementation of performance measurement	128
Figure 23: lack of resources as a potential hindrance for measurement	136
Figure 24: variables for the content analysis of parliamentary questions	144
Figure 25: policy sectors for the analysis of MP's questions	147
Figure 26: an example of the content analysis of parliamentary questions	147
Figure 27: an example of the content analysis of parliamentary questions	148
Figure 28: an example of the content analysis of parliamentary questions	148
Figure 29: supply and demand of indicators	149
Figure 30: indicator classes by policy field	152
Figure 31: quality of supply by policy field	153
Figure 32: supply minus demands	154
Figure 33: potential impact of the use of performance information on degrees of freedom	167
Figure 34: transition table to three functions of performance information	169
Figure 35: implementation strategies	170
Figure 36: effects of the use of performance information	193
Figure 37: central hypothesis of measurement effects: utilization determines effects	201
Figure 38: central hypothesis of measurement effects: breakdown for categories of use	202

Figure 39: overview of the methodology and data	203
Figure 40: linking use and effects	215
Figure 41: effects and uses in the case of the Ministry of the Flemish Community	218
Figure 42: prospective directions for combining theories on organization and information	236

## Tables

Table 1: Four dimensions of performance (Dubnick 2005)	14
Table 2: Rational and constructivist conceptions of performance	15
Table 3 Supply and demand of performance information	32
Table 4: A long list of uses of performance measurement information	38
Table 5: Public Management Styles according to the OECD (1997)	81
Table 6: Summary table of research questions	93
Table 7: A classification of research methods	98
Table 8: Focus of performance measurement studies	102
Table 9: The adoption of performance measurement (based on Beyer and Trice 1982)	124
Table 10: The implementation of performance measurement (based on Beyer and Trice 1982)	125
Table 11: Calculating adoption and implementation from the survey data	127
Table 12: Profiles of performance measurement (PM) implementation and adoption	129
Table 13: Measurability hypothesis	130
Table 14: Political interest hypothesis	132
Table 15: Political interest in the information and activities of twelve sections (M; Moderate, H; high)	133
Table 16: Scale hypothesis	133
Table 17: Street-level discretion hypothesis	135
Table 18: Means hypothesis	136
Table 19: Goal orientation hypothesis	137
Table 20: Frequencies of the occurrence of direct links ( <i>In parliamentary questions with indicators</i> )	150
Table 21: A quantitative assessment of indicator orientation by policy sector	151
Table 22: Eight purposes that public managers have for measuring performance	161
Table 23: Differentiation of implementation mode according to the purpose of performance information.	171
Table 24: Definition of the measurement object.	172
Table 25: Differentiation of criteria for the definition of the measurement object according to the purpose of performance information.	174
Table 26: Differentiation of the choice of the indicators according to the purpose of performance information.	175
Table 27: Advantages and disadvantages of different data sources	176
Table 28: Differentiation of the data sources according to the purpose of performance information. (-; less useful; +; useful)	177
Table 29: Foundations for targets	179
Table 30: An illustration of the breaking out and aggregation of data.	180
Table 31: Differentiation of the choice of analysis technique according to the purpose of performance information (-; less useful, +; more useful).	182
Table 32: Target groups of performance reporting	183

Table 33: Formats of performance reporting	183
Table 34: Differentiation of the reporting format according to the purpose of performance information.	184
Table 35: Differentiation of the quality dimensions according to the purpose of performance information. (* moderately important; ** highly important; *** critical)	187
Table 36: Use related to design: summary table	188
Table 37: Indications from survey data for the initial user profiles	205
Table 38: Selection of the cases based on policy sector, parent department and initial user profile	206
Table 39: Eight purposes that public managers have for measuring performance (Behn 2003) - items in the semi-structured interviews	207
Table 40: Triangulation of the survey based user profiles by means of semi-structured interviews	209
Table 41: The phrasing of the research question and description of the effects	212
Table 42: The effects of performance measurement: assessment per section, average, count, distribution over effects, and probability ratio per effect	213
Table 43: The motive and the opportunity for goal displacement due to performance measurement: limiting factors of the effects and counteracting strategies	219
Table 44: Performance measurement (PM) profiles	230
Table 45: Weiss' (1979) models of research utilization	240

## PART ONE: A PUBLIC ADMINISTRATION PERSPECTIVE ON PERFORMANCE MEASUREMENT

Part one positions our study of performance measurement in the public sector. First, we provide the context for our study. This introductory chapter is intended to give an idea about what we study and what we do not study. Next, the general framework introduces the vocabulary of performance measurement. Moreover, it looks in more detail at the supply and demand of performance information. Thirdly, the subject is positioned in time. We study demand and supply of performance information throughout history. Fourthly, the literature study positions our research vis-à-vis more recent research on performance measurement in the public sector.



# 1. The context of performance measurement

This is a study about performance measurement in the public sector. Yet, before we commence, we need to identify the context of the study. This chapter intends to give an idea about what we study and what we do not study. First, we characterize performance *measurement* in relation to its parent concept - performance of the public sector (1.1). We continue to argue that increasing performance is the central promise of the public sector in post-war Western societies (1.2). Next, we substantiate the claim that measurement is a central facet of public sector reform, public management and public policy and therefore is worth studying (1.3). Fourthly, we position our study in relation to previous research on performance measurement in Flemish public sector (1.4). We make a distinction between the study of the metrics of performance measurement on the one hand and the study of the organizational embedment of performance measurement on the other. Our study takes the latter perspective. Fifthly, we briefly represent the central research questions (1.5) and describe the structure of the text (1.6).

## 1.1. Performance in the Public Sector<sup>1</sup>

Performance is a synonymic word. Dubnick (2005) asserts *“outside of any specific context, performance can be associated with a range of actions from the simple and mundane act of opening a car door, to the staging of an elaborate reenactment of the Broadway musical “Chicago”. In all these forms, performance stands in distinction from mere “behavior” in implying some degree of intent (p391).”* There are two aspects of performance in the literature: (a) the quality of the actions being performed, and (b) the quality of what has been achieved because of those actions. The attention given to the significance of the two dimensions can be high or low. This allows distinguishing between four types of performance.

---

<sup>1</sup> The delineation of the public sector is not straightforward. There are at least three different ways to discriminate between the public and the private (Dutch Social and Cultural Planning Office 2004). (1) First, according to the legal definition, the public sector includes government organizations and organizations governed by public law. (2) Secondly, according to the financial definition, the public sector includes all organizations that are (mainly) funded by public means. (3) Thirdly, according to the functional definition, the public sector includes all organizations that deal with the main functions of government -irrespective from the funding or the legal form of the supplier. Thus, the public service sector may be a more accurate designation than merely the public sector. The focus of our study is plainly public sector. The organizations we study are within the Ministry of the Flemish Community, which is the public sector in the legal, the financial and the functional sense.

		<i>Focus on Quality of Performance Achievement</i>	
<i>Focus on Quality of Performance Actions</i>		Low	High
	Low	Production (P1)	Results (P3)
	High	Competence (P2)	Productivity (P4)

Table 1: Four dimensions of performance (Dubnick 2005)

The most basic dimension of performance focuses the attention on tasks being carried out by the performing agent (P1). What is the production of government? There is low attention for the quality of either actions or achievements. This conceptualization is relatively neutral in nature. Performance in the public sector is about creating public value just as performance in the private sector is about creating private value (Moore 1998). Public sector economists give some insight into the performance of the public sector, vis-à-vis the private sector (Moesen and Van Rompuy 1997). The public sector has to correct market failures or to adjust distributions (Musgrave 1959). The performances of the public sector may be seen as the provision of public goods, quasi-public goods and merit goods. Public goods are non-excludable and non-rival. The non-excludability means that once the good is provided, others cannot be stopped from consuming it. This prohibits profit-making by firms because free-riders could also benefit from the provision of these goods, without bearing the costs. Non-rivality implies that the consumption by one individual does not hypothecate consumption by others. Therefore, rationing is often not desirable. Efficient government is a public good *par excellence* (Stiglitz 1988). For quasi-public goods, exclusion is possible but difficult and consumption is non-rival up to a specific point. Finally, a merit good is a product or service that the government believes consumers undervalue because of imperfect information. Typical examples are education and health care. The societal benefits of consumption are not included in the individual decisions of consumption. In this study, we take the production of the public sector for granted. We will not discuss whether the provision of goods belongs to the public or the private sector and whether they create public value or not. Our focus is on the representation of performance by measurement.

The other dimensions of the concept 'performance' contain a value judgment. Performance has a quality that can be either high or low. First, when performance is about the quality of the actions, and not as much about the quality of the achievements, performance equals competence (P2). Under the assumption that a highly competent performer will be more likely to generate more and better quality output from an activity most of the time, performance becomes associated with the competence of the performer (Dubnick 2005: p392). High performing public sector organizations are organizations that have superior capacity. Secondly, when performance is about the quality of the achievements and not as much about the quality of the actions, performance equals results (P3). The capacity of the organization is not the focus of this conceptualization. In this case, it is the results that count. Below, we will argue that results may be both the outputs and the outcomes of the public sector. Finally, when performance is conceptualized with the attention for both the actions and the achievements of the



organization, then Dubnick equates performance with productivity<sup>2</sup>. Again, productivity is more than merely the output provision. Productivity in this sense has a broad meaning. It refers to the productive organization, i.e. an organization that has the capacity to perform and converses this capacity into results – outputs and outcome. Performance in this study refers to the last conceptualization. We will study how measurement of both capacity and results is embedded in public organizations.

Since performance implies a value judgment, it may only be performance in the eye of the beholder. This question has led to discussions about the objective or subjective nature of performance and the assessment of performance. The schism in the social indicator movement in the 1960s for instance was between the Scandinavian and the American schools. The bifurcation arose from the conceptualizations of societal performance. The Scandinavians stressed the objectively measurable living conditions while the Americans looked at the subjective perceptions of welfare (Noll 1996; Bulmer 2001). Today, public performance continues to be assessed with both objective and subjective performance measures (Van De Walle 2004; 2005). Trust, confidence and satisfaction in government are seen as good performance.

		<i>The value judgment</i> <i>How do people assess performance?</i>	
		Subjective	Objective
<i>The phenomenon</i> <i>What is performance?</i>	Subjective	Strong interpretative stance	Hybrid B
	Objective	Hybrid A	Strong positivist stance

Table 2: Rational and constructivist conceptions of performance

Theoretically, this is a choice between a interpretative and a positivist stance (Table 2). The strongest interpretative standpoints state that not only people’s representation of performance, but also the entity in itself to which the representation refers, is socially constructed (Goldman 2001). The other side of the spectrum is a positivist approach where performance is a reality and performance assessment is an objective process. The hybrids may see the assessment of performance as a subjective representation of real performance (hybrid A) or as an objective deliberation process of subjectively constructed realities (hybrid B). Hybrid B is not as unrealistic as one might expect. Researchers and evaluators often apply rational research instruments to phenomena that they consider to be subjectively constructed. We do not take a stance a-priori. We are interested in performance as people see it in practice through measurement systems. Typically, measurement systems combine subjective and objective indicators and apply a combination of objective and subjective means of assessing the quality of performance.

---

<sup>2</sup> The more restricted definition of productivity mirrors the definition of efficiency. Productivity is ratio of output over input while the definition of efficiency is input over output (infra).

## 1.2. The promise of performance in the public sector

In post-war western societies, increasing public sector performance has been the main promise. At first, this may appear a whimsical statement - what else would be the promise of the public sector? Yet, there are other values in government besides performance.

Hood (1990) distinguishes between sigma type values, theta type values and lambda type values. Lambda-type values are set to keep the public sector robust and resilient. Government has to keep operating even in adverse 'worst case' conditions and to adapt rapidly in a crisis (p14). Reliability is often an argument for choosing public production instead of private production. Theta-type values intend to keep government fair and honest. Government has to pursue honesty, fairness and mutuality through the prevention of distortion, inequity, bias, and abuse of office (p.13). These values are institutionalized in appeal mechanisms, public reporting requirements and ethical codes. Sigma-type values allege to keep it lean and purposeful - to match resources to defined tasks. Thus, frugality of resource use in relation to given goals is the criterion of success, while failure is counted in terms of instances of avoidable waste and incompetence (p12). The promise of performance aligns with the latter set of values. Yet, the other values remain crucial. Allan Schick (1998) argues that in developing countries, the development of old-fashioned, rule-based instruments has to precede New Public Management style instruments such performance contracting<sup>3</sup>. He sees theta and lambda values as preconditions for performance.

In Western societies, the promise of increasing performance is prevailing. Ingraham (2005) observes that *'for much of the twentieth century - and certainly for the last 25 years - performance has been a siren's song for nations around the world (p.390)'*. The post-war expansion of the welfare state has raised expectations about the role of government. In the 1980s, this expansion was no longer supported (Naschold 1996; Pollitt and Bouckaert 2004). Fiscal stress pressured the public budget and legitimacy crises pressured the politico-administrative system. Many people saw government as the problem rather than the solution (Schiavo Campo and Sundaram 2001). As a response, governments pledged to do more with fewer resources - a government that works better and costs less (U.S. National Performance Review 1993). Government reformed in the name of performance. In particular, in the UK and the USA, this led to cutback management and a reduction of the size of government (Dunleavy, 1986; Thompson 2001). Other countries followed other trajectories. Pollitt and Bouckaert (2004) identify four strategies: to minimize (privatize), to marketize (bringing private sector techniques

---

<sup>3</sup> There are seven doctrinal components of the New Public Management (Hood 1990). (1) letting the managers through hands on professional management, (2) setting explicit standards and measures of performance, (3) reliance on output controls, (4) a shift to disaggregation of the public sector, (5) a shift towards greater competition (6) a stress on private sector style management, and (7) a greater discipline and parsimony in resource use.

and values into government), to modernize (changing public sector techniques and values) and to maintain (using the old techniques more intensely).

The societal demand for a high-performing public sector filters through to the organizational level. As a result, one of the central questions for organizational performance research has been whether management matters (Ingraham, Joyce and Donahue 2003; Boyne and Walker 2005). The answer seems to be positive. Admittedly, organizations are constrained by the socio-economic context and the rule of administration and law. Wilson documented the constraints (financial and legal) in government agencies (1989). Still, a growing number of studies prove that management does indeed matter (e.g. Boyne 2004; Brewer and Selden 2000; Meier and O'Toole 2002). A 2004 seminar at Cardiff University has been taking stock of the determinants of organizational performance. Andrews et al. (2005) found that the pursuit of a representative bureaucracy correlates negatively with the perception of performance in English local government. Yet, having an organizational strategy mitigates this effect. Other contributions prove amongst others the positive impact of frontline supervisory management (Brewer 2005), managerial change (Hill 2005), and the negative impact of goal ambiguity (Han Chun and Rainey 2005).

The centrality of performance in contemporary public administration motivates this study. Yet, this is not an exploration of performance as such. We are looking at the representation of performance with performance information obtained through performance measurement. We argue that because performance information is pivotal in public sector reform, public management and public policy, this focus is justified.

### **1.3. Measurement of performance is a key facet of the contemporary public sector**

#### **1.3.1. Performance information is pivotal in public sector reform**

Performance measurement systems did play a pivotal role in reform initiatives (Sterck and Bouckaert 2003). In particular, the United Kingdom witnessed a boost in indicators by the end of the '80s (Pollitt 1986). Reform initiatives such as the Financial Management Initiative, the Next Steps agenda, and the Citizen's Charter led to the creation of performance indicator systems for most public services, central and local. League tables have been created for amongst other schools and hospitals. In the U.K., performance indicator systems have been used as a tool to increase pressure on the public sector, both in terms of efficiency and effectiveness. The Blair administration did not reverse these reforms.

Continental Europe has not used performance indicators with the same intensity. Performance indicators have been used, but not with as high of a profile as with the U.K. There is considerable

variation between countries (Sterck and Bouckaert 2003). In Germany, the 'New Steering Model' (*das Neues Steuerungsmodell*) stressed the importance of performance indicators. However, the reform has only been applied in some big cities, city-states and Länder (Hendriks and Tops 1999). Nowadays, the reform enthusiasm seems to be over and there is increasing acknowledgement of a reform fatigue (Röber and Löffler 1999). In France, the "Loi Organique Relative aux Lois de Finances (LOLF) introduced a form of performance budgeting (Waintrop 2004, Rochet 2005). In Sweden, performance measures mainly played a role in the steering of agencies. The Swedish public sector is highly decentralized. In Norway, the Management by Objectives and Results system has been widely adopted, albeit after a transformation and translation by the agencies (Laegreid, Roness and Rubecksen 2005). The country with the strongest tradition in performance measurement in continental Europe probably is the Netherlands. The first initiatives were taken in the 1970s and by the 1980s, several local governments implemented NPM like measurement-based reforms (Hendriks and Tops 1999). The first large scale implementations of performance oriented reforms at a central level took place in the 1990s. The series of reforms culminated in 1999 with the 'VBTB' initiative – an outcome-based budget structure.

In Flanders, performance measurement and indicators also played a central role in several reform initiatives (Bouckaert and Auwers 1999a, Bouckaert and Auwers 199b).

- The *Doelmatigheidsanalyse* (1992-1998) (efficiency and effectiveness analysis) aimed at the development of a methodology for the analysis of the inputs, activities, output and effects in relation to predetermined objectives. The project did not surpass the experimental phase. Yet, there was an important 'knowledge creep' (Weiss 1978). The concepts of the *Doelmatigheidsanalyse* became common terms of reference within the Flemish Administration.
- The strategic plan for Flanders (1996) contained policy indicators for 13 policy sectors. Indicators have been included in the policy plans of most of the ministers. Yet, there was some reluctance to make the results widely available.
- Thirdly, the *PIP/PEP* (*personeels evaluatie plan / personeels implementatie plan*) (personnel planning) (1997), purposed to quantify existing functions (packages of service delivery, workload, productivity) and to assess future needs to deal with future demands.

Bouckaert and Peters (2002) argue that performance measurement is the 'Achilles Heel' of many public sector reforms. The availability of performance information is a necessary - not sufficient - condition for the success of many reform initiatives. Yet, often the availability of performance information is assumed. The presence of performance information is one of the most decisive and susceptible aspects of the recent tide of public management reforms. This perspective is an important motivation for the focus on measurement of performance in our thesis. Yet, performance measurement goes beyond public sector reform. It is found in recurring activities in public management and public policy.

### 1.3.2. Performance information plays a central role in public management.

Performance information is not only pivotal in public sector reform. It also plays a role in daily management practice. The Government Performance Project provides some insight in the role of performance information in organizations (Government Performance Project 2003; Ingraham, Joyce and Donahue 2003). The Government Performance Project was a six-year research initiative valuating the management capacity of federal, state and local government entities in the USA. The most visible part of the project was the graded reports of the 50 States (Barrett and Greene 2005). The underlying model of the assessment identified four management subsystems that contribute to management capacity – defined as the potential for performance (p.28). The four subsystems are financial management, human resources management, capital management and information technology management. The GPP identifies two crosscutting levers. First, leadership is the driver. Leaders are able to make informed decisions, to provide guidance and direction, to develop the institution's mission, vision and values, to communicate these to the members and to coordinate organizational components. Secondly, information and a focus on results is the connector. Information connects the management subsystems with each other. It also connects the management system with the outside world through measurement of program delivery and performance.

We support this view on the role for performance information in the management of organizations. Yet, we also acknowledge that performance information is only one connector besides others. The 60-year old Friedrich – Finer debate on accountability systems points to an important addition (Bouckaert 2004). Finer championed a system based on objective accountability. He would support performance information for its integrative potential. Friedrich advocated a system based on professional '*fellowship*' between practitioners. Pride related arguments of these professionals allow for a subjective accountability mechanism derived from their values (p. 462). Bouckaert (2004) asserts that balancing the two positions is the desirable option. This should result in a fair mix of trust and measurement, of integrity and compliance and of subjective and objective approaches (p.463).

Figure 1 gives an indication of performance measurement in organizational management for the Belgian public sector (Van Dooren and Van De Walle 2004a; 2004b). The observations are organizations that submitted a good organizational practice for presentation at the first Quality Conference of the Belgian Public Sector<sup>4</sup>. These good practices are mainly management innovations. The application was judged by an external assessor and an international jury. The external assessors made a judgement on, amongst others, the extent to which the organization sets targets for the good practice and the extent to which the organization is developing performance measurement to assess the performance of the good practice. The assessors gave a score from 1 to 5 on both dimensions.

---

<sup>4</sup> The organizations come from the whole Belgian public sector: local, regional and central government.

The scatter plot for the 64 organizations indicates that target setting and measurement are linked. Moreover, it shows that the three profiles are present. Many organizations are setting targets for their good practices. Almost half of them are also measuring progress towards these targets. There are no organizations that are measuring without setting targets for the expected improvement. Finally, some organizations are not measuring nor setting targets for their improvements. Performance measurement for management appears to be very unevenly implemented in the Belgian public sector.

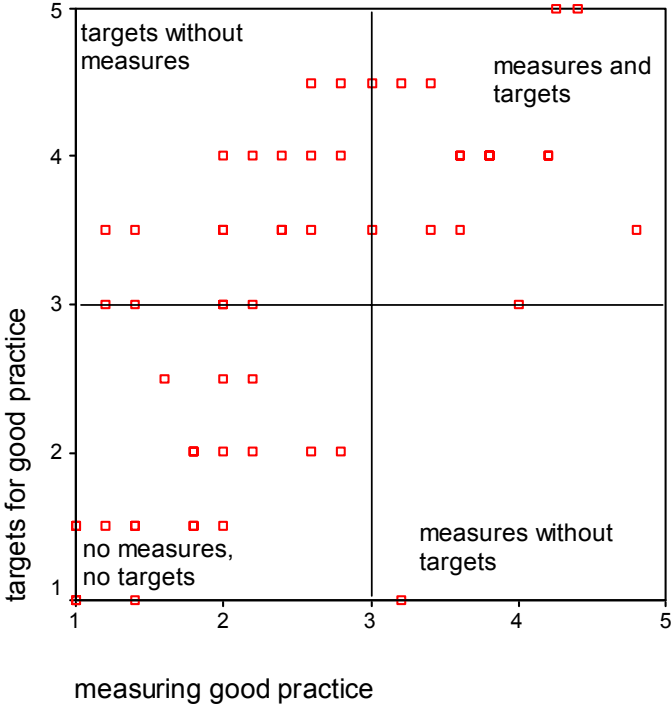


Figure 1: Measurement in organizational management: indications from the Belgian public sector

1.3.3. Performance information plays a central role in public policy cycle.

Performance information also plays a role in public policy. Performance measurement and policy evaluation are adjacent fields. Wholey (1994) sees a role for performance data in the evaluability study that may precede an evaluation. Weiss (1998) points to performance data as a data source for evaluators (see also Newcomer 1997; Wholey 1997; Davies 1999). Some authors go further and advocate an integration of performance measurement and evaluation. McDavid and Hawthorn (2006) assert that performance measurement may be seen as an approach to evaluation. The basic program evaluation tools are also useful for performance measurement. They are complementary evaluation strategies. Yet, some important differences remain.

McDavid and Hawthorn (2006: p.293) point to seven differences. (1) performance measurement systems are ongoing while evaluation is episodic, (2) performance measurement address general issues while program evaluation is issue specific, (3) performance measures are routinized while

evaluation measures are customized for each evaluation (4) performance measurement generally takes attribution for granted while for evaluation it is a central issue, (5) resources for performance measurement are usually part of the organizational infrastructure while resources for evaluation are targeted, (6) managers often play a key role in performance measurement while evaluators and managers are less connected, and (7) the uses of performance information evolve over time while the intended purposes of program evaluation are usually negotiated up front.

Not only evaluators are using information from performance measurement systems. It is infused into the actual policy process. At several points in time, the promise of more objective policy decisions led to substantial measurement efforts. In the 1960s, the promises of the social indicators movement were said to be a 'little short from magnificent' (Smith 1981: p740). Social indicators were to be integrated into a system of social accounting. This system was to reflect the national income accounting system (Gross 1966). Social indicators should also serve as a tool for program evaluation (Rossi and Gilmartin 1980). Moreover, indicators were expected to help set goals and determine priorities. The OECD for Instance chose this social goal approach, which resulted in a series of reports (Carley 1981; OECD 1973,1976). In recent years, the thrust towards evidence-based policies again propagated enlightened decision-making based on what works (Davies, Nutley and Smith 2000; Taylor Fitz-Gibbon 2002). The evidence based policy movement revealed lessons from the social indicator movement. Evidence is seen as embedded in the political and policy rhetoric of the day, and infused in the newly transformed professional ethic of many service professionals (p.11).

Evaluation in the Belgian Public sector is only modestly institutionalized. Varone et al. (2005) distinguish three impediments for institutionalizing evaluation. First, the power of the political parties (the *partitocracy*) works against allowing new 'veto' players (such as evaluators) into the policy process. Moreover, it leads to a disturbance of the stages of the policy process. Policy evaluation has to jump several hurdles before it reaches the decision arena. This will only happen when the evaluations are adopted by the political parties. Secondly, the power of the parliament in Belgium is low and declining. Unless the Executive consents, evaluations have little chance of being put on the parliamentary agenda. A third barrier is the Belgian version of federalism. The federation and the regions and communities have shared and sometimes competing competences. Yet, evaluations often cut across these institutional boundaries. As a result, implementation of the conclusions becomes more complex. Although evaluation is not institutionalized in the sense that there are formal organizations and that there is an epistemic evaluation community, performance indicators are used in the policy planning.

In particular the Flemish public sector relies on performance indicators for policy planning. Flemish public administration and agencies collect and reports significant sets of policy relevant information that give insight into the performance of policies (Bouckaert, Van Dooren and Sterck, 2003). Important sets concern educational and environmental achievement and employment figures. Also, indicators

are included in the policy plans of ministers. Although this is not always the result of structured, ideal-typical deliberation process, many quantitative indications of policy preferences are included. Only 13.6% of the indicators also set a target (Conings, Sterck, Van Dooren and Bouckaert 2005).

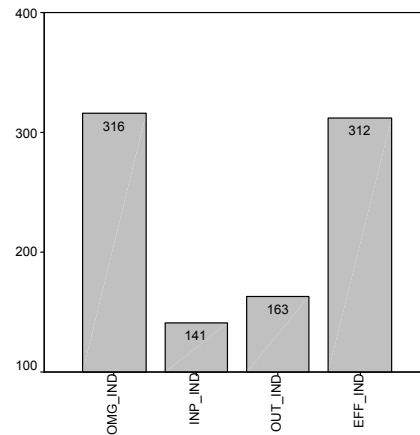


Figure 2: Count of the number of indicators in the policy plans of the ministers of the Flemish government (Conings, Sterck, Van Dooren and Bouckaert 2005)

#### 1.4. Performance measurement research in Flanders

Finally, we position this study vis-à-vis past performance measurement research in Flanders. Until now, two research fields have performed public sector performance measurement research; research in economics, mainly dealing with the metrics of performance measurement, and action/policy oriented research in Public Administration<sup>5</sup>.

In Belgium, there is a tradition of performance measurement research in economics. The main concern is the generation of good metrics for measuring performance. The development and application of non-parametric frontier methods is the more longstanding tradition. These methods are particularly suitable for the public sector. Because of the relatively weak knowledge of the production function in the public sector, parametric estimation is difficult (Moesen, 1992). The development of the Free Disposal Hull method was an original contribution in this context (Deprins, Simar and Tulkens 1984). FDH was developed in the context of a comparison of labor inefficiency in post offices. They found that 48% of the post offices were fully efficient. They argue with regard to the method that *'while only few managers – whether private or public - would have any reason to impose that the operations of their firms be organized so as to fit the graph of an a priori selected function (or the boundary of a*

---

<sup>5</sup> Public Administration in upper case refers to the academic discipline, while public administration in lower case refers to the practice (Brans 2003).



*necessarily convex set*), most of them would agree with the statement that using less input for the same or a larger output is better than using more (p.264)'. Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH) analysis has been applied on several public services afterwards. Bouckaert studied fire services and civil registry offices (1992; 1993). Van Den Eeckhout et al. (1993) and De Borger et al. (1994) made a comparison of Belgian municipalities. Tulkens (1993) applied the nonparametric tests to three dissimilar sectors: retail banking, courts and urban transit systems. Recently, Moesen and Persoon (2002) performed a study of tax offices. Besides the nonparametric frontier methodology, more recent research focuses on the development of synthetic performance indicators (Cherchye, Moesen and Van Puyenbroeck 2004). The aim is to get to a subsidiarity respecting synthetic indicator for social inclusion. This project is of particular relevance in the context of the Open Method of Coordination of the European Union.

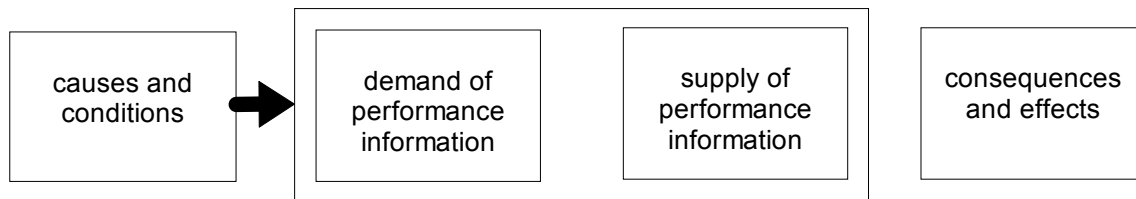
Secondly, there is a research tradition in performance measurement research in Public Administration. Still, the research has predominantly been action research and applied research. The action research aimed at supporting the implementation of the *Doelmatigheidsanalyse* in the Ministry of the Flemish Community (Bouckaert, Van Reeth, Auwers, Verhoest 1998). The research revolved around several pilot studies and was published into a handbook. Secondly, the applied research was predominantly an international comparative (Bouckaert, Hoet and Ulens 2000; Depeuter, Bouckaert and Van Dooren 2003; Sterck and Bouckaert 2003). The purpose of these research projects was to draw lessons from international practices. These studies did not center around the metrics of performance measurement. They have a Public Administration focus. The main research questions concern the institutional (most notably political) environment, the organizational structure and culture, and the behavioral implications for decision-makers and implementers.

Our study has a Public Administration perspective and aspires to go beyond applied and action research. Our units of analysis are public sector organizations that are (not) measuring their performance and are (not) using this information for policy and management purposes. We will research the reasons why organizations are measuring, the tuning of supply to demand, and the effects of measurement on organizations. We regard these research questions as complementary to the research tradition that studies the metrics of performance measurement.

## **1.5. Research Questions**

In part two of the study, the empirical study, we will study the following research questions on performance measurement. We also position the research questions in the overall framework of the thesis. The arrow indicates the overall relation we will research.

## Why are organizations measuring performance?

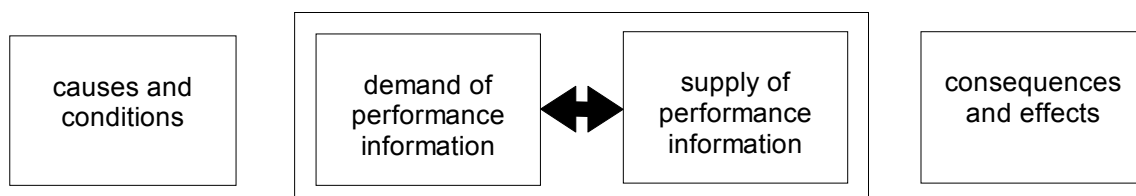


Although performance measurement is the Achilles heel of many public sector reforms, there is seldom an explicit policy to obtain performance measurement information (Bouckaert and Peters 2002). Availability of performance measurement is assumed, and reforms are built on this unsteady foundation. Insight into the organizational and contextual factors that facilitate or impede performance measurement is crucial for developing a performance measurement policy. Governments may be tempted to design a one-size-fits-all policy, often based on best practices. However, differences between organizations may be considerable and should be taken into account. This section attempts to empirically establish some of the factors that explain differences between organizations.

Six research issues will be studied.

1. The measurability issue; is the type of output a determinant?
2. The political interest issue; leads political interest or political apathy to measurement?
3. The scale issue; do large organizations measure more than small ones?
4. The street level discretion issue; does street level discretion play a role?
5. The means issue; is it mainly a matter of resources?
6. The goal orientation – issue; is the coupling with goals of main importance?

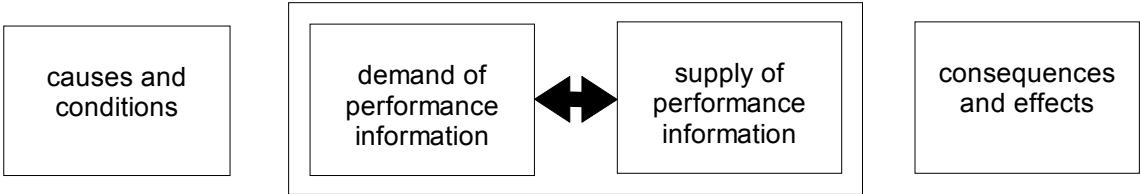
## Does administrative supply meet political demand?



Civil servants often criticize politicians for not using performance information. Politicians complain that reliable performance information is not available. Politicians and civil servants complain that their own policy sector is hard to measure, if not immeasurable, compared to other sectors. This section searches for an empirical insight into the administrative supply and political demand of performance information.

Three sets of research questions will be studied. First, to what extent is there a supply and demand of indicators? To what extent do supply and demand meet? What is the quantity of the (mis)match? Second, what is the quality of the supply? What are the motivations, if any, for not providing the demanded information? An additional assessment on quality is the explanation for not providing the demanded indicators. Thirdly, are the differences between policy sectors significant and if so, which policy sectors measure more?

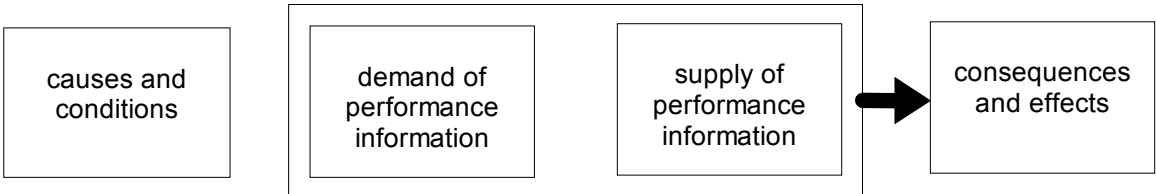
**What are the system requirements for different purposes of performance information?**



One of the most important research questions in performance measurement is “*how to design measurement systems that are ‘fit for use’?*”. However, the question should be supplemented with “*fit for which use?*”. Behn’s article “*Why Measure Performance? Different Purposes Require Different Measures*” (2003), relates uses to the choice of the indicators. Here, we will relate uses to other steps in the measurement process such as for instance target setting, analysis techniques and reporting

Two research questions will be addressed. First, what are the uses of performance information? Use is the independent variable in this research setting. Therefore, we need a classification of uses that goes beyond the techniques and instruments listed on page 36. Secondly, how does use relate to design? How to design a measurement system that is fit for its purpose?

**What are the effects of performance information?**



Measurement of performance is not neutral. Since the Hawthorne experiments, we know that the act of measurement in itself is influencing performance. By measuring performance of organizations, the organizational behavior is affected. Several authors addressed this issue. This resulted in laundry lists of potential unanticipated effects of performance measurement. This section will first describe the most

important effects from the literature. Next, we will look search for a common denominator. We will then establish which are the empirically most important effects. Finally, we will examine the hypothesis that the effects are a consequence of a specific profile of use. This viewpoint challenges the tendency of effect-studies to view performance measurement as a one-dimensional concept.

Four research questions will be studied. What are the effects of performance measurement in the literature? What is the common denominator in these effects? What are empirically the most important effects? Which effects arise from which use?

## **1.6. Structure of the text**

First, the object of this study will be defined. We study the supply and demand of performance information from a public administration perspective. Next, we will look at the history of performance measurement. Supply and demand of performance information evaluated throughout history. Then, we will study the recent literature in order to position our study vis-à-vis recent studies. We will also introduce the middle range approach to theory. In the fifth chapter, the methodology of the study will be described. After that, the corpus of the thesis are four empirical issues on performance measurement. Finally, some overall conclusions are drawn.

## 2. General Framework: what is performance measurement?

Many see performance measurement systems as a black box – an indivisible unit of analysis. Hatry (2002) considers the conceptual entanglement of performance measurement and performance management as one of the main fallacies in performance measurement theory and practice. Thus we have to get inside the black box. Performance measurement is multi-faceted. This chapter defines the many meanings of performance measurement in the practice of public sector organizations. This anatomy of performance measurement is an essential part of our study. We will argue in chapter 4 in favor of a middle range approach to Public Administration. We also will argue that the Public Administration discipline bases its identity on shared subjects of study then on shared theoretical lenses through which researchers are looking at reality. Therefore, we need a general framework that describes the subject of study.

This chapter serves as a general, analytical framework of performance measurement in the public sector. First, it is analytical since it identifies different parts of performance measurement systems. The general framework thus enables an insight in the performance measurement system. It provides a description of the subject of study, i.e. performance measurement systems. As mentioned earlier, it allows for a dissection of the black box. What are the constituents of a performance measurement system? Secondly, it is a framework. It is a steppingstone for the empirical research questions. Obviously, we do not study all the research questions one could possibly identify within the framework. However, the research questions we do study, are included in the framework.

First, the vocabulary of performance measurement is clarified. These concepts are familiar to many. They are part of the mindset of both academics and practitioners when looking at government. However, despite (or due to) its popularity, the concepts are plagued by dysfunctional differences in interpretation (Bouckaert 2004; Pollitt and Bouckaert 2004). Section 2.1 gives our interpretation of the main concepts in performance measurement. Next, section 2.2 describes the supply and demand analogy for performance measurement. On the one hand, the use of performance information triggers a demand. On the other hand, the production process supplies information that should be 'fit for use'.

## 2.1. The Vocabulary of Performance Measurement

Figure 3 represents the logic model that underpins performance measurement. It is a model of the role of an organization or a program in the socio-economic environment (Pollitt and Bouckaert 2004). Policy evaluators generally use the same logic (McDavid 2006).

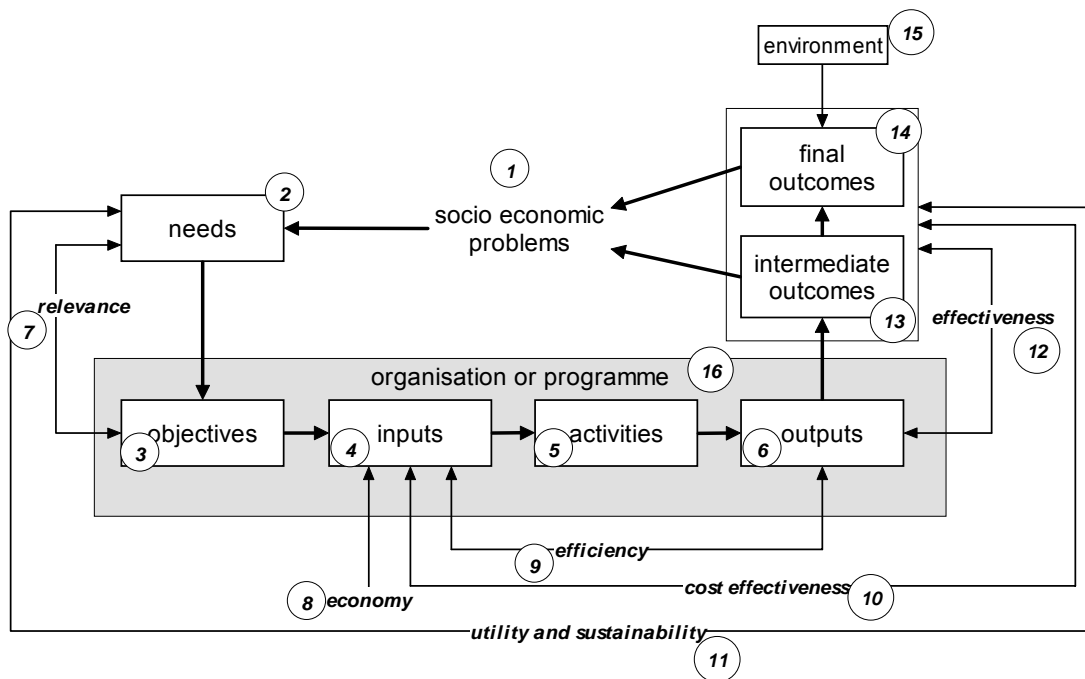


Figure 3: The input-output model

The starting point is the socio-economic situation. Socio-economic issues (1) invoke a need (2) for action by the public sector. This role is in an ideal-typical situation attributed to politicians<sup>6</sup>. Certainly, not only politicians translate issues to problems and problems to policies (Parsons 1995). Yet, the political system will filter issues and determine priorities. The confrontation of the objectives of a policy (3) with a need allows assessing the relevance (7) of the pursued policies. Next, resources (4) are allocated in order to stage activities (5) that yield outputs (6). Economy (8) is the ratio of a monetary input over another input (e.g. the cost of a computer). The ratio of the input over the outputs is the efficiency (9). Some authors define productivity as the inverse ratio of efficiency: output over input. (Hatry 1999).

<sup>6</sup> The ideal typical distinction between policy-politics on the one hand and management-administration on the other hand will seldom be found in reality. The administration often is involved in policy decisions. Politicians often interfere with very concrete cases. Svava (1985) gives a good overview of the different configurations of the role of politicians and administrators in mission, policy, administration and management.

The outputs are expected to have effects on society. These effects can be intermediate (usually in the short term) (13) or final outcomes (usually in the long term) (14). The final outcomes in particular are influenced by the environment (15) on which the organization or the program has a limited or no impact. The ratio of output over effect is the effectiveness (12). The ratio of the input over the effects is the cost-effectiveness (10). The outcomes of a program or an organization have to address the needs of society. The confrontation of needs and outcomes allows assessing the sustainability and utility (11) of the program or the organization. Only the objectives, input, activities and outputs of the organization or the program (16) are under direct control. The other factors depend largely on the impulses of society.

As an example, we apply the model to the issue of traffic casualties. Politicians formulate the need to reduce the number of casualties in traffic. Typically, several interest groups and the (perceived) pressure from their constituency influence them. Flemish politicians use comparisons of several countries to call for a stricter policy. The objective is to reduce the number of casualties to a number comparable to other developed countries. In order to attain this goal, government will use (financial and other) resources to build cycle tracks, to reconstruct crossroads, to install speed traps, etcetera. The outputs are the kilometers of new tracks, the constructed new crossroads and the number of controlled vehicles. To this point, government has a good grip on the chain of events. The decisive test however is the outcome in society. In the short run, it may be that more children cycle to school and that fewer drivers violate the speed limits. This is the intermediate effect. Government however wants to reduce the numbers of casualties. Do the immediate outcomes lead to the final outcomes? Undoubtedly, environmental factors will interfere. For instance, a reduction in the number of casualties may be the result of bad weather conditions. In cold and rainy weather, there are usually less cyclists and pedestrians, and therefore less potential victims. Yet, driving conditions are worse, and therefore there is a higher chance of accidents.

We based the input-output model on Pollitt and Bouckaert (2004; p.106). This is one of the more extensive models. The model above (a) changes one ratio, (b) adds two ratios, (c) omits two labels and (d) adds one factor to the original model. The adjustments are made in order to align the model with other authoritative models. The three alternative input output models are from Hatry (1999: p.24), ASPA's Centre for Accountability and Performance (2000: p.20), and Poister (2003: p.37).

(a). We define one ratio differently. Pollitt and Bouckaert (2004) define effectiveness as a relation between objectives and outcomes (intermediate and/or final outcomes). Other models however define effectiveness as the relation between output and outcomes. Two points argue in favor of the latter approach. First, if effectiveness relates only to the objectives, only intended effects are taken into account. Outputs usually have unintended positive and negative side effects. Cost-benefit analyses refer to these effects as externalities (Ammons 2002). These externalities may be included in the assessment of the effectiveness of a program. Secondly, objectives may be attained merely by

changes in the environment without any intervention of a program or an organization. The objective to reduce unemployment for instance depends largely on economic growth. Employment programs will only have a second order impact. In this case, the effectiveness is not the effectiveness of the program or the organization.

(b) The unemployment example makes the case for the addition of the 'environment' field. Poister (2003) for instance includes these environmental influences in his generic program logic model. These external influences may impede or facilitate success. Often, these external influences are client characteristics of the program or the organization. A school for instance does not always choose its intake. Nonetheless, the socio-economic status of the pupils heavily influences the educational outcome. These factors can be particularly helpful in making sense of performance data.

(c) Two additional ratios are incorporated. First, we included economy. Economy is about the input of an organization, and therefore may not appear of the foremost relevance. However, economic management may release extra resources for policy. Moreover, many public managers are primarily concerned with economy, for instance in facility management. Economy may be an important performance indicator for supporting services (ICT, infrastructure) in an organization. Secondly, the ratio of input over effect is included. This is cost-effectiveness. It may be seen as the societal value for money. Notwithstanding its conceptual appeal, the measure is usually hard to calculate because of the impact of external influences on outcomes.

(d) We omit two labels. We only talk about intermediate and final outcomes. The terms impact and results are somewhat confusing. The notion result is sometimes used for output, but also for outcomes in general (both final and intermediate). The Government Performance Project at the Maxwell School of Public Policy in Syracuse (USA) comprises both output and outcome under the chapter managing for results (GPP 2002). Impact too has different meanings (Hatry 1999). Some analysts use the term impact to refer to data that estimate the extent to which the program or organization actually caused particular outcomes. Others use impact to denote the societal outcomes as opposed to the outcomes for the individual. In order to avoid confusion, this study uses the term intermediate and final outcomes.

**Performance measurement in this study is the quantitative representation through measurement of the quality or quantity of input, output, and/or outcome of organizations or programs in its societal context.**

This representation may consist of different types of indicators.



We base the typology of indicators on the input-output model (Bouckaert and Van Dooren 2003). A distinction can be made between single indicators and ratio-indicators. Ratio-indicators are combinations of single indicators. In most instances, they yield additional information compared to single indicators. Yet, some information (i.e. the absolute values) is lost by calculating ratios.

### → Single indicators

Indicators on input	What goes into the system? Which resources are used?
Indicators on output	Which products and services are delivered?
Indicators on intermediate outcomes	What are the direct consequences of the output?
Indicators in final outcomes	What are the final outcomes of the output?
Indicators on the environment	What are the contextual factors that influence the output?

### → Ratio indicators

Economy	input/input
Efficiency	input/output
Productivity	output/input
Effectiveness	output/outcome (intermediate or final)
Cost-effectiveness	input/outcome (intermediate or final)

## 2.2. The Object of Study: Supply and Demand of Performance Information

The object of the study is the supply and demand of performance information in organizations. We take a Public Administration perspective on performance measurement. We will look at the role of performance information in organizations. We study why and how performance information is produced, how it is used and with which effects. First, we describe the supply and demand approach. Next, we briefly look at the constituent parts of supply and demand.

### 2.2.1. Supply and demand, causes and effects

The object of the study is on the one hand the production process of performance information and on the other the use performance information. Information may be considered as goods and services (Machlup 1972). One party provides information; while another party consumes it. For instance, the government supplies a ranking of schools in a league table and the press and the citizens consume. Or, oversight bodies demand information of agencies to be included in performance contracts. A very important issue is who supplies the performance information. Is it the department or - as is often the case - the agency itself?

Two issues somewhat confuse a clear-cut demand and supply interpretation. First, the supplier and the consumer are often the same organization. For instance, an organization might supply indicators that are to be used in a quality model and thus largely for its own organizational policies. Nonetheless, the distinction between supply and demand remains useful, albeit an intra-organizational supply and demand. The study of intra-organizational supply and demand requires closer insight into the organization. A second issue is whether the provision of indicators is demand-driven or supply-driven. The conventional position is demand-driven. The need for information determines the supply. For instance, *New Public Management* reforms have created a substantial demand for indicators. Influential authors proposed a results oriented government (e.g. Barzelay 1992; Osborne and Gaebler 1993). Regularly, the demand for information is enacted in legislation. This is particularly the case at the central government level and less at the local level. The Government Performance and Results Act in the USA is an example of legislation that triggered a demand for performance information (Van Reeth 2002). Yet, the supply of performance information does not need to be demand driven. It may be the other way around. Organizations may supply information and thereby create a demand. The viewing figures for television in Flanders are an example. Both public and private television channels stressed importance of viewing figures for years - alternately in accordance with how successful the channels had been. Nowadays, these figures are paramount and difficult to ignore. Innes (1990) demonstrates how the availability of the GDP measure explains its success. In private sector, the creation of a supply (mainly through publicity) is a legitimate and deliberate strategy. In the public sector, the same applies. Even so, it appears to be less of conscious strategy.

The theoretical and practical agenda should be a search for a better match between supply and demand. This implies that more measurement is not necessarily better measurement. Progress in measurement means the provision of the right quantity and quality of information, fit for a specific purpose. We illustrate this by positioning supply and demand on a 2 x 2 matrix (Table 3).

		<i>demand</i>	
		Low	High
<i>supply</i>	Low	Position A: No indicators	Position B: Demand frustration
	High	Position C: Supply frustration	Position D: Supply and demand

Table 3 Supply and demand of performance information

**Position A:** There is no supply of and no demand for information. Progress would imply advancement from position A to position D. However, also a movement to positions B and C is possible. This position implies that there is no quantitative evidence for a given problem. At most, there is a qualitative, ideological or normative identification of the problem.

**Position B:** There is no or low supply, but there is demand for information (i.e. undersupply). There is an intention to use information. However, it is hampered by the lack of performance information and measurement. This results in a demand frustration zone. For instance, politicians nowadays often ask for data on citizens' trust and client satisfaction. However, these concepts are often vague and difficult to measure (Luhmann 1998, Lewis and Weigert, 1985).

**Position C:** There is no or low demand, but there is supply of information (i.e. oversupply). Performance measurement is developed but the information is not used. This results in a supply frustration zone. This box encompasses early warning information, which is ignored. Another example is the Planning Programming Budgeting System, which provided enormous amounts of data for which there was no demand (Schick, 1966, Wildawsky 1969). At several points in history, information was supplied without a clear demand for it (infra)<sup>7</sup>. The supplier intended to influence policy makers. In some instances, the information became institutionalized (cf. Innes 1990) on the GDP). Demand followed supply.

**Position D:** There is demand for and supply of information. Examples of high supply and demand are the statistics on economic growth, unemployment, etc. These indicators are highly institutionalized. For instance the economic indicators are largely uncontested. The social-indicator movement was the first movement to agitate against the 'economic philistinism' (Bauer 1966). However, until now, the traditional economic indicators such as the GDP still have a much greater audience compared to quality of life indicators.

Most of the elements and functions of performance measurement may be subsumed under the supply (i.e. the production process -*box 3 in Figure 4*) or demand categories (i.e. the uses/consumption - *box 2 in Figure 4*)<sup>8</sup>. Supply and demand is the core of performance measurement research and practice. For practitioners, the crucial question is how to develop the production process of performance measurement in such a way that it yields performance information that is fit for use. Additionally, two other important categories of research questions may be formulated with supply and/or demand as either the dependent variable or the independent variable. First, the performance measurement system may be the dependent variable that is explained by organizational, cultural, structural and/or institutional factors. What are the causes and conditions for performance measurement (*box 1 in Figure 4*)? Secondly, the performance measurement system may be the independent variable

---

<sup>7</sup> Chapter 1 explores the history of measurement in government in more detail. One conclusion from the historical study which is of relevance here is that performance information increasingly found its way into government. The first initiatives were taken by external actors which we today would call interest groups.

<sup>8</sup> Chapter 4 reviews the performance measurement literature. The research questions are categorized within the framework of figure 3.

explaining effects in the organization or the environment (*box 4 in Figure 4*). A well-documented research line in the latter respect is the study of perverse effects of performance measurement.

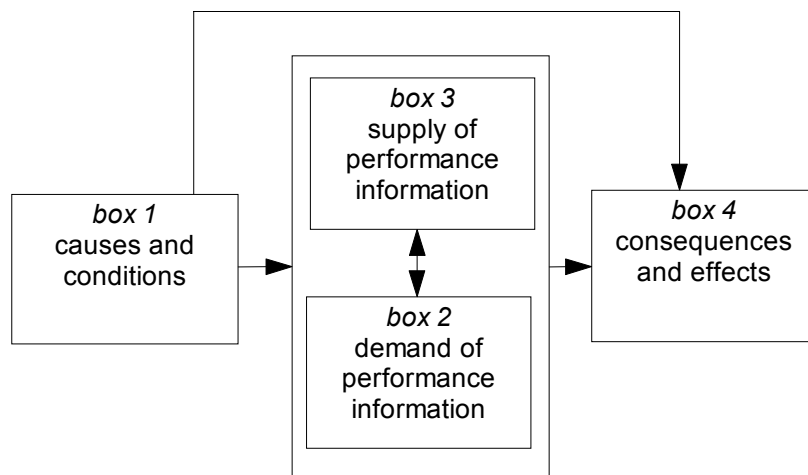


Figure 4: Research questions regarding performance measurement

Hence, four categories that cover the majority of the research variables in performance measurement research may be identified (i.e. the boxes of *in Figure 4*);

- variables describing the **causes and conditions for performance measurement**,
- variables describing **how the performance information is used** (demand/consumption side),
- variables describing **how the information production process advances** (supply side),
- variables describing the **effects of the introduction of performance measurement** on the organization and its environment?

The provision and use of performance information is an observable fact that we will try to explain. The provision and use of performance information may be observed at three different levels. At a micro level, organizations are measuring their performance. Performance based management for instance is a micro issue. The meso level is about the political processes and policies (economic, labor market, and education policies) in policy sectors. Some policy sectors may have a more positive disposition towards performance measurement than other sectors. Finally, the macro level is the government wide level. Macro indicators assess the performance of the government as a whole. The performance measurement issues studied in this research deal with micro and meso supply and demand.

The production process of performance information and the use of that performance information is the object of the study. In the section 2.2.2 and 2.2.3, respectively the supply side and the demand side are discussed.

## 2.2.2. The production process: supply of performance information

Figure 5 depicts the performance measurement process in six steps. It is an ideal typical representation of the production process of performance information. As an ideal-type, the process is a rational one. It is based on the one hand on practice oriented literature on performance measurement. Harry Hatry's book (1999) for instance has been an important source of inspiration. He too looks at performance measurement as a process that yields information that then can be used in a number of ways. On the other hand, it is based on unstructured interviews in the Ministry of the Flemish Community. The interviews assured us of the fact that the model has practical value. Thus, although the ideal-type will not materialize fully fledged, practitioners will recognize the components. This is a necessary condition for doing the research. Therefore it is an important heuristic scheme. When we will introduce middle range theorizing in the empirical research, we will mitigate this rationalistic stance. We follow Simon's (1976) argument that rationality is bounded by less rational factors.

The organization can make several choices in each phase. These are design parameters that lead to different measurement configurations. Chapter 8 provides a detailed description of the design parameters, related to the use of performance information. Here, we give a brief overview of the process.

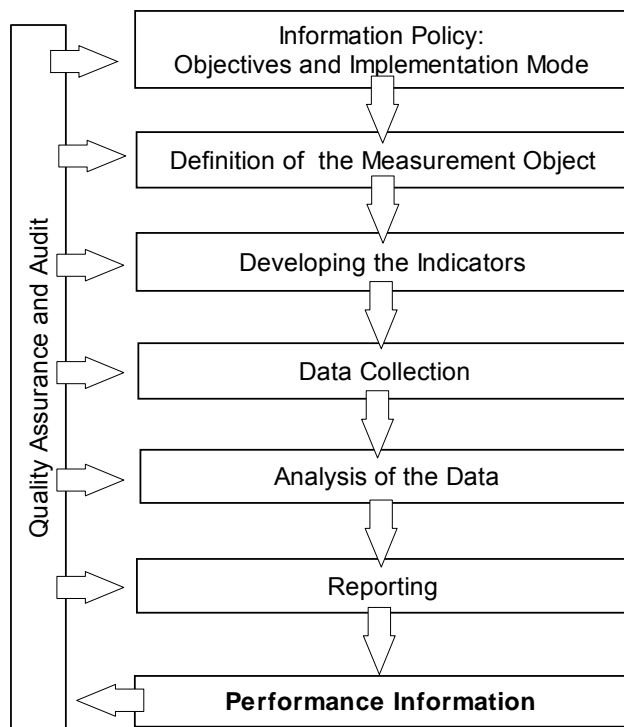


Figure 5: The production process of performance information (supply)

The first step is the development of an information policy. The objectives of the measurement efforts as the implementation mode need to be established. The objectives are a forecast of the intended use. As we will argue in chapter 8, the use will have to determine the design parameters of a measurement process. In other words, it is an articulation of the demand. The second step is the definition of the measurement object. The organization has to decide what is going to be measured and thus, what is not going to be measured. Thirdly, the indicators need to be developed. They should be good representations of the measurement object<sup>9</sup>. It is important to realize that an indicator does not equal the underlying reality<sup>10</sup>. Fourthly, the organization has to collect data. Indicators without data are empty boxes. Yet, regularly, organizations formulate indicators and put them 'on hold'. This is justifiable as a developmental strategy. Fifthly, the analysis of the data has as a purpose to get from data to information. The metrics of performance measurement are mainly applied in this step. Three methods may be identified; the confrontation of the result with a norm or a target, causal analysis, and aggregation and disaggregation. The last step is the reporting phase. The right reports need to be drafted for the dissimilar target groups. The production process leads to performance information with a certain quantity and quality. The quality assurance of performance information is a separate point of interest and may have a bearing on all the phases in the production process of performance information.

### 2.2.3. Uses of performance information

There may be several uses of the information. Several authors discussed the uses of performance information (see list on the next page). Note that these are the uses as they can be found in organizational practice. This explains the length of the list. For research purpose, we will define categories of use. Chapter 8 will provide and motivate a categorization of uses of performance information. Here we confine ourselves to presenting a long list of uses of performance information. We listed the uses with nearly the exact phrasing of the authors. Therefore, differences between uses

---

<sup>9</sup> Some authors distinguish between an indicator and a measure. A measure in this definition is a precise representation of reality while an indicator only gives an indication. We do not make this distinction. The degree of reality of measurement is not an absolute and objective categorization. Rather, the degree should be assessed on a scale ranging from zero to total representation. It would be hard to identify the point on the scale where an indicator becomes a measure. Moreover, this assessment is a subjective one and will depend on who is the evaluator.

<sup>10</sup> This fallacy refers to what Etzioni (1967) calls concept reduction. He gives the example of the IQ test. Often, the scores on this test are equated with intelligence. Yet, the IQ test may as well represent a particular kind of intelligence - i.e. logical reasoning. Synthetic thinking as well as emotional intelligence are not tested. Another example is the criticism on the GDP. It points precisely to the fact that is a measure of economic growth, but not of societal welfare (Dowrick and Quiggin 1998). A typical example is the investment in private security.

are nuanced. The purpose of presenting the long list of uses here is to demonstrate what the use of performance information (the demand side) may denote in organizational practice. We reviewed ten performance measurement texts and listed the different proposed uses of performance information.

1. Mayston (1985) discussed the role of non-profit performance indicators in the public sector. He was predominantly concerned with nationalized industries.
2. Osborne and Gaebler's (1993) New Public Management text on Reinventing Government prescribes several ways of putting performance measures to work (p155).
3. The Government Accounting and Standards Board (GASB 1998) reported about a survey of USA local and state governments on the use of performance information.
4. Wang and Berman (2000) conducted surveys in USA counties.
5. The USA's General Accounting Office (2000) surveyed the USA federal agencies.
6. The OECD (2003) has data on the use of performance measurement in the OECD member states.
7. Hatry (1999) proposes ten uses in a performance measurement text that is mainly conceived as a manual for practitioners.
8. A similar text by Broom et al (2000) of ASPA's Centre for Accountability and Performance also has a list of uses.
9. Poister (2003) describes the uses of performance information
10. The Governance Performance Project (2002) reviews the management capacity of states and cities in the USA and incorporates the use of performance measures in its assessment.

	No of Occurrences	Mayston 1985	Osborne & Gaebler 1993	GASB 1997	Hatry 1999	Wang & Berman 2000	Broom et al 2000	GAO 2000	OECD 2003	GPP 2003	Poister 2003
1. allocation of resources	5				X	X		X	X	X	
2. changing work processes / more efficiency					X	X		X	X	X	
3. formulation and monitoring of licensed or contracted privatized services	4	X			X			X		X	
4. rewarding staff / monetary incentives / performance pay			X		X			X		X	
5. strategic planning				X	X		X				X
6. communication with the public to build trust	3			X	X	X					
7. reporting and monitoring				X		X					X
8. accountability to elected officials					X		X			X	
9. accountability to the public					X		X			X	
10. results based budgeting: budget documents			X	X	X						

Although they add to the GDP, few would argue that they are an indication of societal welfare.

	No of Occurrences	Mayston 1985	Osborne & Gaebler 1993	GASB 1997	Hatry 1999	Wang & Berman 2000	Broom et al 2000	GAO 2000	OECD 2003	GPP 2003	Poister 2003
11. results based budgeting: justify budget requests				X	X					X	
12. motivation rewards for groups, organizations			X				X			X	
13. evaluation of outcomes and effectiveness		X		X		X					
14. reducing duplicative services / delivery alternatives (incl. privatization)				X		X	X				
15. adopting new program approaches / changing strategies				X			X			X	
16. setting program priorities							X	X	X		
17. communication with the legislature and the legislative staff				X		X					
18. cost saving	2	X		X							
19. performance budgeting: no direct link			X				X				
20. setting individual job expectations/ staff performance plans								X	X		
21. cost benefit analysis		X			x						
22. trigger for further investigation and action		X			X						
23. enable consumers to make informed choices		X									
24. improving responsiveness to customers				X							
25. creditor reporting							X				
26. grantor reporting							X				
27. output budgeting: pay per output (p * q)			X								
28. outcome budgeting: pay per outcome			X								
29. changing appropriation levels				X							
30. performance budgeting: alongside budget figures							X				
31. cost accounting											X
32. performance auditing											X
33. capital management	1										X
34. managerial incentive schemes		X									
35. management by objectives			X								
36. staff motivation/ non monetary incentives					X						
37. strategic HRM											X
38. clarifying objectives		X									
39. Quality Models (TQM)			X								
40. sanctioning prolonged low performance					X						
41. allocating discretionary funds to high performance agencies or programs					X						
42. communication between managers						X					
43. organizational development							X				
44. coordination of activities internally or externally								X			

Table 4: A long list of uses of performance measurement information



The longlist of performance based practices in organizations gives an indication of what the use of performance measurement in organizations may signify. The most frequently mentioned uses are the allocation of resources, changing work processes and increasing efficiency, the formulation and monitoring of licensed or contracted privatized services, rewarding staff / monetary incentives / performance pay, and strategic planning.

The presentation of the long list of uses at this point only has an informative purpose. It is used to determine the subject of study, i.e. a public administration approach to performance measurement and the use of performance information. However, a long list of 44 uses is not a useful tool for studying performance measurement. In order to study the use of performance measurement, we will have to classify the uses in a more limited set. A classification is proposed in chapter 9. We will arrive at three categories of use; use for research and learning, use for internal management, and use for accountability purposes.



## 3. A measurement history of government

### 3.1. A historical argument for studying performance measurement

In the early nineties, government allegedly needed reinvention (Osborne and Gaebler 1992). Bureaucracy was said to have gone bankrupt in the course of the last decades (p.12). Before, during the World Wars, the Depression and the New Deal, the bureaucratic model was believed to work well. It provided stability and a basic sense of fairness and equity. It also provided jobs and basic, no-frills, one-size-fits-all services people needed during the industrial era (p14). In the 1970s and 1980s, the model became subjected to great pressure. *“We now live in an era of breathtaking change. We live in a global marketplace, which puts enormous pressure on our economic institutions. We live in an information society, in which people get access to information almost as fast as leaders do. We live in a knowledge-based economy, in which educated workers bridle at commands and demand autonomy. We live in an age of niche markets, in which customers have become accustomed to high quality and extensive choice (p15).”* The old bureaucracies were said to be maladjusted to this brave new world. Therefore, we witnessed a paradigm shift -a shift from a bureaucratic paradigm to a post-bureaucratic paradigm, where the entrepreneurial spirit is about (Barzelay 1992: p118). Moreover, the paradigm shift is said to be a global phenomenon, at least in the developed world (Osborne and Gaebler 1992: p328). Publications of the Public Management (PUMA) section of the Organization for Economic Cooperation and Development (OECD) subscribed to this viewpoint enthusiastically, and promoted the entrepreneurial spirit amongst its member states (OECD 1995; OECD 1996).

Osborne and Gaebler (1992) wrote the right text at the right time. The publication fitted neatly into the reform agenda of the first Clinton administration. The publication heavily influenced Vice-President Al Gore’s *National Performance Review (NPR)* (US National Performance Review 1993; Moe 2003; p21). The reformulation of the role of government as catalyst for private initiatives (*‘steering, not rowing (p25)’*) allowed for a rigorous saving program without abandoning the goals of the public sector. Savings were substantial. The NPR website<sup>11</sup> claims that the NPR *“reduced the size of federal civilian workforce by 426,200 positions between January 1993 and September 2000. The government workforce was the smallest it had been since the Eisenhower Administration.”* It is also claimed that *“action on more than two-thirds of NPR recommendations resulted in savings of more than \$136 billion.”* The focus was mainly on the machinery of government rather than on program cuts.<sup>12</sup> The aim

---

<sup>11</sup> <http://govinfo.library.unt.edu/npr/howeare/appendixf.html> (accessed 15/11/2005)

<sup>12</sup> For a thorough evaluation of the NPR, see Vol. 60, Issue 6 of the *Public Administration Review*. Kettl (2000) identifies devolution and globalization as the two main challenges for governance. Thomson (2000) analyses the

of the NPR is 'not simply to weed the federal garden; it is to create a regimen that will keep the garden free of weeds (National Performance Review, reprinted in Schafritz et al.; p561)'. The popularity of the concept of governance fitted well into this reasoning. Governance can be seen as "steering capacity", or some mechanism of providing coherent direction to society and constraining, prescribing and enabling the provision of publicly supported goods (Peters 1998; Heinrich and Lynn; 2000). Thus, government plays one role (amongst other actors) in providing governance. According to Osborne and Gaebler, government as an instrument for providing governance is outdated, although the provision of collective goods is not necessarily contested. It is a matter of means, not ends. This allowed for a compromise between the Democratic program and the disastrous state of the US federal budget in the early nineties.

First, was there really a paradigm shift? Many doubt it (see e.g. Williams 2000). Academics are particularly critical about the New Public Management (Pollitt and Bouckaert 2004: p138). Although it may not have been a new paradigm, management reform has not all been windy rhetoric (Pollitt, 2000: p195). There are results; downsizing, efficiency gains, more flexibility and user friendliness. Secondly, was it really a global trend? Again, there are serious doubts. Pollitt (2001) finds convergence between countries in reform strategies a useful myth. He uses Brunson's (2002) framework to distinguish between four types of convergence with a differing degree of 'purity'<sup>13</sup>. Christensen and Laegreid consent when they describe the transformative processes that would make any global movement to a unique local translation. Central New Public Management ideas are transformed when they face a local context (Christensen and Laegreid 2002; Christensen, Laegreid and Wise 2002).

The critics reduced NPM from a paradigm shift to an Anglo-Saxon management fad (Goodsell 2004). At least, the NPM reforms appear to be a victim of what Sieber (1981) calls '*over commitment (p.161)*'. He states that administrative reforms often defeat themselves by initial hype that produces later disillusion. In any case, the post NPM era seems to have been arrived. Kickert (2003) for instance reviews the Dutch 'Tilburg model'. Tilburg is a medium sized city in the Netherlands that was acknowledged as a NPM champion. At the same time that the city had to hire an external consultant to organize the reception of delegations, the city itself was already moving away from the model (p. 395) (see also Hendriks and Tops (2003)). Performance measurement was pivotal for NPM reform

---

results of the NPR reforms and seeks explanations for its failures. Williams (2000) points to some inconsistencies and flaws in the NPM literature. Kearny, Feldman and Scavo (2000) report on the perception of city managers. Denhardt and Denhardt (2000) formulate a synthesis between the old public management and the new public management; the new public service model.

<sup>13</sup> Type A convergence is convergence in results. This is the purest form. Type B is convergence in actions, although the results may not converge. Type C is convergence in decisions, while the actions and the outcomes may diverge. Finally, the lowest form is convergence in talk, not in implementation.

initiatives. The catchphrase of the NPR - *from red tape to results*- plainly encompasses a results-orientation. Christopher Hood (1991) distinguishes nine doctrinal components of the New Public Management. Two out of nine components relate directly to performance measurement. First, there is the doctrine of explicit standards and measures of performance. This implies the definition of goals, targets indicators of success, preferably expressed in quantitative terms. The justification is that accountability requires a clear statement of goals and efficiency requires a “hard look” at objectives. Secondly, Hood identifies a greater emphasis on output controls. Resource allocation and rewards have to be linked to measured performance. This is justified by the need to stress results rather than procedures. The tight coupling of performance measurement with the NPM makes performance measurement suspect for what went wrong in the application of NPM.

The critics of New Public Management risk to drag performance measurement along. NPM is believed to be bad, performance measurement is a central part of NPM, and so performance measurement too is believed to be bad. Although performance measurement is a part of NPM, it does not equal it. Seven out of nine of Hood’s doctrinal components of NPM are not directly related to the measurement of performance. In 1969, Aaron Wildavsky argued for rescuing Policy Analysis from PPBS. He feared that “*policy analysis will be rejected along with its particular manifestation in PPBS (p190)*”.<sup>14</sup> The same may be said of the relation between performance measurement and NPM reforms. By assimilating performance measurement with its NPM applications, the potential of performance measurement may be wiped out. The potential of performance measurement lies in its ability for an organization to improve decisions, accountability and performance in itself (Epstein 1992; Hatry 1999). Therefore, it may be useful to look at measurement of government performance well beyond the last decades of NPM reforms.

### **3.2. Research questions for the historical study**

Measurement in government is not a recent phenomenon. Williams (2002;2003;2004) for instance studied thoroughly measurement in the early 20<sup>th</sup> century in the USA. This section has a different approach. We attempt to give an overview of the main movements in measurement history of government –at the expense of a more profound insight into the single movements. However, this birds-eye approach to history may give insight in issues, which may not so easily be addressed by a study of a single movement. We will look at the following issues. Is there real change or are we constantly reinventing the wheel? What is the dynamic of rise and decline of measurement movements? Is measurement in government really an Anglo-Saxon dictum? Does measurement have private sector or public sector roots? What are the contingencies with the societal context? What are

---

<sup>14</sup> One of critics of PPBS and the rescuers of policy analysis (as mentioned by Wildavsky), William Gorham, supported the development of performance measurement at the Urban Institute.

the carriers of ideas across time and space? Is there an evolution in supply and demand? By giving a descriptive account of the measurement history, we suggest some answers.

### 3.3. Overview

The purpose of the study of the measurement history is to contextualize measurement here and now. We look at movements that provide information on the output and outcomes of public sector activities. We use secondary sources to describe the evolution of measurement in government. A history must start somewhere, but history has no beginning (Kendall 1960). It is therefore always possible 'to pursue the roots of a subject down to its slenderest fibrils' (p447). Kendall argues that history of statistics begins with the Political Arithmetic around 1660. This will be the starting point for this study .

The objects of the study are movements of thought and practice that attempted to measure government. The concept of 'movement' is analogous to the sociological term 'social movement'. Unlike other forms of organization, movements are informally organized (Jary and Jary 1999).

The overview of the history of performance measurement comes with a warning. As we noted above, it is an overview and therefore it can only be general idea and an outline of the history. It has the advantages and disadvantages of a distant view. The advantage is the more encompassing mapping of the field. The disadvantage is the potentially cursory reading of history. Several supplementations are conceivable. First, more movements may included. This overview identifies 14 movements that had an important impact on the quantification of government. However, every movement has its antecedents and offspring. It would be audacious to claim comprehensiveness. Secondly, movements may be omitted and/or replaced. Other researchers may use other sources and come to different conclusions regarding the importance of the movements. Thirdly and related to the second point, this study does not weigh the importance of movements. It would nonetheless be interesting to determine real impact on government. However, this would require in-depth document analysis of primary sources, which goes beyond the scope of the study at hand. Fourthly, studies may be more in-depth. At least three approaches are conceivable. First, in-depth studies can focus on a single movement. The studies of the New York Bureau of Municipal Research are an example (Williams 2003; Schachter 2004). Secondly, studies may explore in depth a specific time segment and thirdly studies may focus on a geographical circumscription.

The chronology of movements that contributed to the development of performance measurement is represented in Figure 6. This figure does not have the ambition to be the alpha and omega of the intellectual history of performance measurement. This is our reading of the measurement history. It sketches several important movements in developing measurement. The description of the movements covers four topics in four paragraphs.

- a. Definition: What is the movement about and what are its main manifestations?
- b. Context: What was the societal context of the movement?
- c. Example: What is a typical example of a study?
- d. Relevance: What is the main contribution to the development of performance measurement?

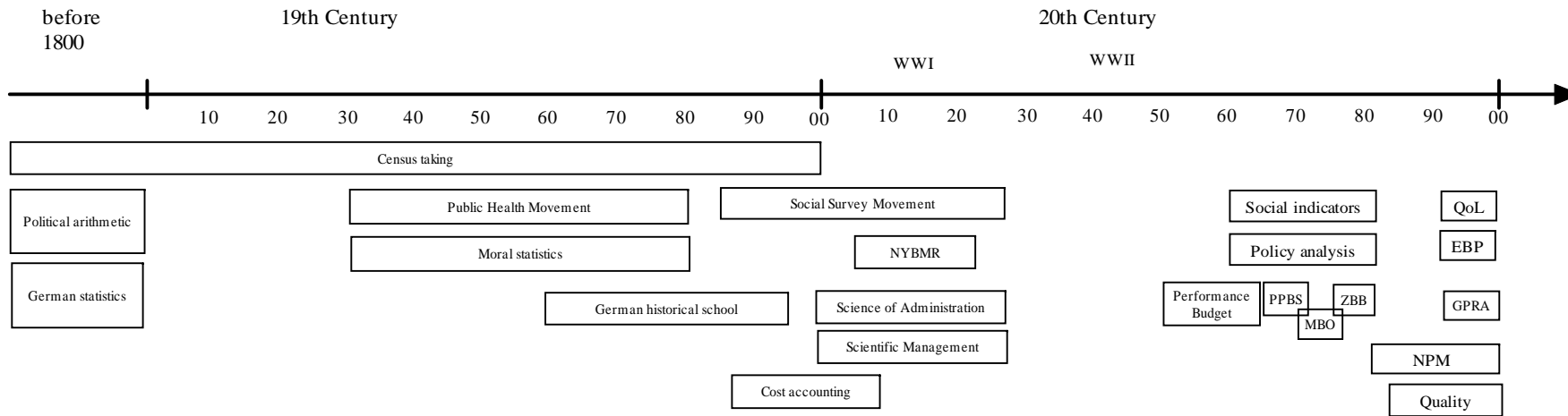


Figure 6: A chronology of movements

NYBMR: New York Bureau of Municipal Research

PPBS: Planning Programming Budgeting System

MBO: Management By Objectives

ZBB: Zero Based Budgeting

GPRA: Government Performance and Results Act

NPM: New Public Management

QoL: Quality of Life

EBP: Evidence Based Policy



### 3.4. Political arithmetic

#### Definition

*“Instead of using only comparative and superlative Words, and intellectual Arguments, I have taken the course (as a Specimen of the Political Arithmetick I have long aimed at) to express my self in Terms of Number, Weight, or Measure (William Petty, Discourse on Political Arithmetic, 1690)”*. This statement is emblematic for the Political Arithmetic whose interest was *“the Art of Reasoning by Figures upon things Related to Government (Charles Davenant, quoted in Innes de Neuville, 1975: p11).”* The political arithmeticians provided practical methods for resolving concrete problems. All or most of William Petty’s writings were prompted by the practical problems of his time and country - problems of taxation, of money, of the policy of international trade (Schumpeter 1972: p.213). Their activities involved three important stages: keeping written records, scrutinizing and assembling them according to a predetermined grid; and interpreting them in terms of numbers, weights and measures (Desrosières 1998).

#### Context

The political arithmetic developed in Britain in the 17<sup>th</sup> century. It was a favorable time for the development of science. By 1660, both parliamentarians and royalist turned to science and technology for its potential economic and social benefits (Jardine 2002). The leading figures of the political arithmetic however, William Petty<sup>15</sup>, Charles Davenant<sup>16</sup> and John Graunt<sup>17</sup>, were not academicians.

---

<sup>15</sup> William Petty (1623-1687) was a self-made man - physician, surgeon, mathematician, theoretical engineer, Member of Parliament, public servant and businessman (Schumpeter 1972). He studied in Leiden and Caen and afterwards in Oxford where he became professor of Anatomy. From this he secured the lucrative post of physician-general to Cromwell's forces in Ireland. Here he changed direction and gained fame for his 'Down Survey' of Ireland. Petty surveyed the entire country in fourteen months, but instead of using skilled surveyors, he completed the project using the unemployed - and cheap - soldiery, and made considerable personal profit. In 1661, he became Sir William Petty and the following year he published his most famous work, *Treatise of Taxes and Contributions*, which discussed the role of the state in the economy. (BBCi history, historic Figures [www.bbc.co.uk/history](http://www.bbc.co.uk/history)).

<sup>16</sup> Charles Davenant (1656-1714) was a public servant and Member of Parliament. As an economist, he takes up an eclectic position, recommending governmental restrictions on colonial commerce as strongly as he advocates freedom of exchange at home (Desrosières 1998).

<sup>17</sup> John Graunt (1620-1674) was a prosperous merchant until his business was destroyed in the London fire of 1666. While still active as a merchant, he began to study the death records that had been kept by the London parishes since 1532. He was mainly known for his studies on the vital statistics.

They were experts with a precise field of competence who suggest techniques to those in power while trying to convince them that, in order to realize their intentions, they must first go through him. A new social role took shape. This could only happen in conditions in which the state became part of society, and not its totality. The liberal concept of the state influenced in still another way the reliance on indirect methods and calculations of the political arithmetic. In 1753, a plan to take a census was denounced by the Whig party as “*utterly ruining the last freedoms of the English people*” (quoted in Desrosières 1998: p24). Calls for a census were resisted in fear of new taxes. Eventually, the first census in Britain was taken in 1801 in order to address the need for food after a disastrous harvest in 1800 (Bulmer, Bales and Sklar 1991). This was half a century after Sweden (1749) and a century after Iceland (1703). The lack of encompassing data forced researchers to develop methods for calculation and estimation.

### Example

The most famous and controversial technique was the population multiplier (Desrosières 1998). Note that population growth was considered the main performance indicator for government. The technique is typical for the work of the political arithmeticians. The challenge was to estimate the total population of a country, taking into account that one could not conduct a census. Parish registers, however, provided the numbers of births. The method consisted of taking a census in a limited number of parishes. Then they calculated the relationships between the births and the population and extrapolated this relation to the whole of the country. This technique, an ancestor of random sampling, was vigorously attacked in the nineteenth century and until the beginning of the twentieth century, exhaustive censuses were preferred.

### Relevance

The political arithmetic provided the tools for quantified assessments and applied them to government. Unlike the German statistics that were merely descriptive in nature, they wanted to explain and to compare. Their data was used in the emergent life insurance sector and in the population debate that was spurred by the mercantilists<sup>18</sup>. A large population was considered important for augmenting the power of the state. Therefore, population figures were the most important outcome measure for

---

His book was titled *Natural and Political Observations mentioned in a following index, and made upon the Bills of Mortality With reference to the Government, Religion, Trade, Growth, Ayre, diseases, and the several Changes of the said City.* (Encyclopaedia Britannica 2002)

<sup>18</sup> Mercantilism was the economic theory and practice common in Europe from the 16th to the 18th century that promoted governmental regulation of a nation's economy for the purpose of augmenting state power at the expense

obtaining the public goal: increasing the wealth of the nation. Although the innovative nature of the Political Arithmetic was important, it was not an extensive movement in quantitative terms.

### 3.5. German University Statistics

#### Definition

It was Gottfried Achenwall (1719-1772) who first introduced the term 'statistics' (Hecht, 1977). However, the meaning differed from contemporary statistics. By the term, he meant a comprehensive description of the social, political, and economic features of a state. The German statistics offered the official a framework for organizing the multiform branches of knowledge available for a particular state. The system would make facts easier to remember, easier to teach, and easier for men in government to use (Lazarsfeld 1961). The program of the courses by the founding father of the German statistics, Hermann Conring (1606-1681), followed Aristotle's *methodus quattuor causarum*<sup>19</sup> (Hoock 1977). The material cause describes the territory and its population. The formal cause entails law, habits and the rights of the prince. The efficient cause deals with the administration and the judicial system. The final cause has to do with the goals of the state; population growth, safety and development. Throughout the 18<sup>th</sup> century, the statistics remained highly descriptive. Illustrative for the narrative nature of the movement is its resistance against constructing cross tables for comparing countries. Cross tabulation would reduce the described objects and make them lose singularity. The partisans saw the benefits of comparison. The opponents labeled these statistics as 'vulgar statistics', finding the information superficial and incomplete.

#### Context

The societal context in Germany was fundamentally different from Britain (Lazarsfeld 1961). While in Britain, industrialization had spread, Germany was still rural and poverty-stricken. The Peace of Westphalia (1648) ended the Thirty Years' War (1618–48)<sup>20</sup> and left a country in a state of collapse

---

of rival national powers. The power of the state was based on natural resources as well as on human resources. Trade was seen as a zero sum game.

<sup>19</sup> Aristotle's teleology gave four causes. A full explanation of anything must consider not only the material, the formal, and the efficient causes, but also the final cause—the purpose for which the thing exists or was produced.

<sup>20</sup> In the Thirty Year's war, the Roman Catholic emperor and other Catholic princes and states fought against a network of Protestant towns and principalities and their foreign supporters. Germany was devastated and lost a large part of its population, and the empire's borders were greatly reduced. Under the terms of the Peace of Westphalia, Germany's numerous feudal princes gained virtually full sovereignty. (encyclopaedia Britannica)

with nearly three hundred rival microstates. German universities were heavily dependent on the states. Professors remained largely unpaid and the student-life was at an all-time low. No middle class existed, no intellectual centre, no national aristocracy, which might have supported the work of artists and scientists (Lazarsfeld 1961). The princes of the states were all confronted with identical problems. All legal disputes over problems of territory, marriage and succession had to be solved referring to case laws and archival studies. This situation gave authority those who described and catalogued rather than those with innovative ideas. The weakness of the microstates and their need for self-definition led to what Désrosières (1998) terms “*a cognitive patchwork* (p20)”.

### Example

A typical study in the German tradition was published in one of the first issues of the Journal of the Statistical Society of London (1839)<sup>21</sup>. The title is “*An Account of the Recent Progress and Present Extent of Manufactures in Prussia, and of the Trade of the Prussian Commercial Union in Manufactured Goods*”. It provided data on the import and export of cotton, the livestock, the number of machines for spinning wool, the employment (constantly and occasionally), and so forth. The different categories are carefully subdivided. For instance, the production of linen is broken out into the production of grey packing linen and sail cloth, raw unfinished linen, bleached, dried or printed linen cloth, ticking, table linen and toweling, etcetera. These analyses yielded detailed accounts of a country. However, contrary to the political arithmeticians, they did not pursue causal insight between variables.

### Relevance

The impact of the German university statistics on the development of quantification goes mostly unacknowledged (Schumpeter 1972; Désrosières 1998). The political arithmetic is seen as the main thrust towards quantification. However, Lazarsfeld (1961) does not agree. The Germans had acquired international audience in the second half of the 18<sup>th</sup> century. At that time, they were probably as influential as the political arithmeticians were. In fact, even in modern social sciences, the development of nomenclature is an honorable pursuit. In the accounting literature too, the Cameralist tradition - which was closely related to the German statistics - has an important place (Forrester 2004). The cameralist accounting was the origin of the single entry bookkeeping. The cameralists furthered the development of budgeting. Finally, the German university statisticians were highly intertwined with government. Schumpeter (1972) described them as consultant administrators.

---

<sup>21</sup> Initially, the Royal Statistical Society in Britain took the German approach to statistics, rather than that of the political arithmeticians (Schumpeter 1972).

### 3.6. Public Health Movement (vital statistics)

#### Definition

The next step in the quantification of '*government at work*' is found in Britain. The administrative sphere and the sphere of social reformers, scholarly societies and university professors were distinct, but interacted intensively. Quantitative research became an instrument for influencing policies, in particular about the living conditions of the poor. The Statistical Society of London, a central institution at the time, centered its discussions on health and social conditions and took an ameliorist approach. The society organized annual conferences, sought to influence Parliament to legislate on social issues, and relied heavily on the influence of its prestigious membership to achieve reform (Bulmer, Bales and Sklar 1991: p.12). The movement was epidemiological and anti-contagionist on the medical front, and environmentalist and reformist at the social level (Désrosières 1998: p.170). Towards the end of the century, the movement was marginalized due to two evolutions. First, bacteriological discoveries revealed the causes and direct treatments of diseases. Statistical epidemiology was less needed. Secondly, the increasingly popular theory of social Darwinism stated that philanthropic measures for relieving the poor favored the reproduction of the most inept and hindered natural selection. Figures about individual occupation were thought to be more relevant than geographic data.

#### Context

The societal context was one of high poverty and mortality due to the rapid industrialization. Half of Manchester children died before their fifth birthday. A laborer in Liverpool claimed a life expectancy of 15 years (Richman 2003). This led to controversies about how to improve social conditions and to combat epidemic diseases (Bulmer Bales and Sklar 1991). On the one hand, there was the harsh approach advocated by Edwin Chadwick -father of the Poor Law. In 1834, the New Poor Law was voted, which forbade relief to the unemployed. Aid to the unemployed had to be organized in workhouses, where the conditions had to be more severe than the conditions of the poorest workers<sup>22</sup>. Actually, they were prison-like institutions. By 1839, the 15,000 parishes of Britain had been grouped into 600 unions while 350 workhouses had been built. The administration of the Poor Law led to the erection of the Registrar General's Office headed by William Farr (1842-1879). On the other hand, there was the Anti-Poor Law movement, Robert Owen and the Factory Movement, and the Chartism.

---

<sup>22</sup> Workhouses were within the responsibility of a central commission and not under direct ministerial or parliamentary control. It had wide powers to establish efficient local administrative units, to supervise the work of locally elected guardians and to prescribe the qualifications of local officials. It was allowed to make regulations for the general administration of relief.

The soft reforms they proposed were not entirely inspired by altruism. It was important to neutralize growing discontent. Unemployed workers formed armed gangs and riots were rife (Richman 2003)<sup>23</sup>.

### Example

A typical example is a study by William Farr on a cholera epidemic in 1849 in Britain (Farr 1852). He studies the “*influence of locality; particularly of elevation of habitation on the diseases and characters of men (p.155).*” He discovered that the majority of deaths occurred in a minority of districts. Moreover, he found that the Cholera was three times more fatal on the coast than in the interior, in particular in the big seaports. Farr attributes the mortality of the cholera to the elevation of the habitat. He searches for reasons. Lower cities receive the wastewater from higher places, and thus all the ‘*extent of organic matter in the air we breathe and in the water (p.163).*’ The practical consequence he derived from his study is that although elevation of habitation, with purity of air and purity of water, does not prevent the cause of cholera, it reduces its effects to insignificance (p164). Farr finished his study with an international comparison with, amongst others, the deltas of the Ganges, the Mississippi and the Nile, the West coast of Africa, Rome, Constantinople and Syria. By collecting and publishing this data, Farr created a market for his products. The publications of Farr and the Registrar General’s office became increasingly authoritative. Mortality rates became pertinent indicators of municipal policy (Désrosières 1998: p168). Farr developed this report during the 1850s. He calculated the average mortality of the healthiest 63 districts and set this rate as a goal for all the others. The British Public Health movement influenced the French scene. The ‘*Annales d’hygiène publique et de médecine legale*’ are a landmark in the development of the French vital statistics (Lecuyer 1977)

### Relevance

The public health movement truly was a next step in the quantification. Firstly, the methods became more sophisticated. Herewith, the tradition of the political arithmeticians was upheld. However, this methodological development was influenced by the moral statisticians on the continent, in particular Adolphe Quetelet. William Farr for instance studied in Paris, and was a great supporter of Quetelet (FitzPatrick 1960). In contrast to the political arithmeticians, they had plenty of data at their disposal, in particular census data. The analyses therefore had more to offer. Finally, there was link with

---

<sup>23</sup> Around 1850, the study of the conditions of poverty and criminality was driven forward by the studies on family consumption (Innes DeNeuville 1975). The founding fathers were Edouard Ducpétiaux and Frederic Leplay. Ducpétiaux was one of the first scientists that made a link between social conditions and criminality (Bockstaele et al 2004). Contrary to Quetelet who saw statistics as a neutral graduator, Ducpétiaux found that statistics needed to serve the social sciences. Statistics had to discover fundamental

policymakers, but not in the same way as with the German statistical movement. The latter were highly dependent on the states. It was not critical research. The Public Health Movement was distinct and independent from the state, highly critical and still influential. A blueprint of the modern think tanks.

### 3.7. Moral statistics

#### Definition

In the 18<sup>th</sup> century, the *quantity* of the population was the most important attribute. A large number of inhabitants were considered as a resource for the state. After the 1830s, a big population was rather seen as a threat. A qualitative and good population was the goal. The moral statistics wanted to measure this quality of the population (Höjer 2001). The key personality of the moral statistics movement was *Adolphe Quetelet*, a Belgian statistician. He was searching for the “laws” of the social world (Lottin 1912; Lazarsfeld 1961). The emphasis was on crime statistics. However, data on e.g. suicide, drunkenness and duels was also included.

#### Context

Moral statistics developed in an era of revolution and restoration. The French Revolution of 1830 triggered a revolution in the Kingdom of the Netherlands and led to the independence of the Kingdom of Belgium. This unrest intensified Quetelet's interest in social affairs. A revolution is according to Quetelet a disequilibrium that is caused by a dysfunctional government. A wise government knows how to deal with the fluxes that cause such disequilibria. One can understand the extent of civilization of a nation by the way it is ‘*doing its revolutions*’. The detection and understanding of the social laws and their application of these laws by government, should enable society “*d’effectuer les réformes de la manière la plus avantageuse (Quetelet, quoted in Perrot 1977)*” - to economize on unnecessary revolutions.

#### Example

The most celebrated work of Quetelet was his ‘*Essai de physique sociale (1935)*’ - Essay on social physics (Stigler 2000). The first three parts of the work described natality and mortality, the growth and the moral characteristics of human beings. The fourth section introduced the theory of the average man (Bockstaele, Cerulus and Vanpaemel 2004). The average was a necessary standard to evaluate

---

inequities and disequilibria. Leplay further studied the family budget in case studies (Lazarsfeld 1961).

deviations in character or health. The average was also considered a model of harmony. The important figures from history came close to the current ideal type for their respective periods. The average man was an abstraction that described the population, not the individual. Moreover, the margins of the population were important. These boundaries indicated the feasible changes and improvements of the average man. An important research theme of Quetelet and the moral statisticians was the *“penchant au crime”* - the inclination to crime. They proved that on average people at different ages have differing tendencies to commit a crime (Lottin 1912).

## Relevance

The relevant points of moral statisticians (with Adolphe Quetelet as their foremost representative) for the quantification of policies and management are twofold. First, they introduced very strongly the idea of a target, i.e. the average. Quetelet wrote that *“l’homme moyen est donc aussi le type de santé et de beauté (quoted in Perrot 1977)”* (The average man is thus the healthy and beautiful type). The underlying philosophy is one of conformation to the middle and condemnation the extremes. Statistics had to lead the statesmen where the average was the guiding light. By this, the numbers were a representation of the condition of society. The pursuit of the average confirmed the contemporary power structure. The efforts by Quetelet to disseminate his science internationally were the second contribution. He was building intellectual institutions (Stigler 2000; p.51). He started a series of conferences between 1853 and 1878. Hundred and fifty scientists from 26 countries attended the first conference in Brussels. The Belgian King was present (Lottin 1912). Other congresses followed, amongst others in London (1860), Berlin (1863), Saint Petersburg (1872). The conferences were the ancestors of the International Statistical Society. The strong ties between scientists were established. This international research arena facilitated the flow of ideas between the statistical societies of many countries (Willcox 1934).

### **3.8. German historical school**

#### Definition

In the 19<sup>th</sup> century, statistical bureaus were established in the German states (Desrosières 1998). Prussia had the most important bureau. However, the other German states -Saxony, Württemberg, and Bavaria - had own bureaus. In 1871, an additional imperial service was created, but the regional bureaus remained independent. In 1934, they were absorbed into the unified statistical office of the Nazi-state. The bureaus provided political, historical and geographical descriptions of a given region. The directors often were university professors who taught the science of the state. Between 1860 and



1862, Ernst Engel was chief of the Prussian bureau. He was innovative in managing his bureau. He contributed significantly to the study of family budgets<sup>24</sup>. Although he was a prominent figure, Engel failed in building a politico-scientific network under Bismarck's authoritarian ruling. Thereafter, Engel founded the *Verein für Sozialpolitik*, which became an influential institution in the scientific world and is known until today in economics. The German historical school's method was inductive. It relied on empirical monographs with historical and statistical data. The main criticism on their work was about the gap between data and recommendations. According to the German historical school, recommendations should flow directly out of historical accounts. No intermediate analysis is sought out. The opponents were the English and Austrian schools that followed an abstract and deductive approach<sup>25</sup>.

## Context

The societal context in Germany altered halfway in the 19<sup>th</sup> century. Industrialization took a firm hold and Germany caught up with England and France. A working class was formed. The working class became organized, with trade unions and a Social Democratic Party. The first compulsory social insurance programs on a national scale were established under Bismarck: health insurance in 1883, workmen's compensation in 1884, and old age and invalidity pensions in 1889. The aim of the reforms was inspired by a fear of the Marxist revolutionary forces. This is reflected in the mission statement of the *Verein für Sozialpolitik*. The *Verein* wanted "*to raise, educate and reconcile the lower classes on the basis of the existing order*" as Gustav Schmoller, the chairman of the *Verein* for many years (1890 - 1917), put it<sup>26</sup>. Due to this stance, they were labeled 'Katheder Sozialisten' - Socialists of the Chair.

## Example

An important study was a survey on the agricultural workers of East Prussia by Max Weber<sup>27</sup> (Désrosières 1998). National identity at the time was frangible. The unification of Germany was recent. Moreover, industrialization caused considerable migration within Germany from Northeast Prussia to the Southwest and the Rhine. Russian and Polish agricultural workers filled the vacant jobs. Weber wanted to describe the new relationships in economic terms and to evaluate the impact on social and national cohesion. The issue was how Germany should evolve given the social upheavals caused by

---

<sup>24</sup> His most important contribution is the known as Engel's law: the proportion of food consumption in a given budget decreases when income increases.

<sup>25</sup> The controversy resulted in the 'Methodenstreid' between the German and the Austrian School.

<sup>26</sup> <http://www.socialpolitik.org>

<sup>27</sup> Weber's early research of which this survey is an example, was within the historical tradition. Later in his career, he developed a distinct school.

migration and industrialization. There were two surveys: One for landowners with factual questions on the numbers of salaried persons, the remuneration, the social characteristics of the workers, the types of labor contracts, and the accessibility of schools and libraries. The second survey addressed teachers, pastors and public officials who should know the values and opinions of the rural communities. The final report contained nearly 900 pages. Weber's goal and conclusions of the survey were mainly political. He recommended a small independent agriculture instead of the vast capitalistic estates that were ran from Berlin by the *Junkers*.

## Relevance

The German historical school preferred the descriptive and inductive approach to science above the abstract and deductive one. In the early 20<sup>th</sup> century, this approach was overruled by the emergence of sociology and econometrics. The movement may seem to have been a dead end. Nonetheless, it was noteworthy for several reasons. First, the relentless will to dig into society yielded an invaluable extension of our factual knowledge (Schumpeter 1972; p803). This serves as a reminder that data and analysis are two sides of the same coin. This is particularly the case when quantification should result in practicable knowledge and concrete policy recommendations. Secondly, many French, English and American scholars studied in and were influenced by German historicism<sup>28</sup>. Thorstein Veblen, John Commons and Westley Mitchell built on the German historical tradition by stressing the importance of the social framework that was in turn shaped by historical and cultural forces (Scott 1995). Thirdly, the offspring of the German Historical School was particularly talented. The youngest generation of scholars of the Verein included amongst others Spiethoff<sup>29</sup>, Tonniës<sup>30</sup> and most importantly Max Weber whose work on bureaucracy laid the foundations for the administrative sciences in Europe.

## 3.9. Census

### Definition

Censuses were important throughout the history of the quantification of government, in particular before the development and acceptance of sampling techniques. The modern idea of a population census as a complete enumeration of all the people and their important characteristics for purposes of understanding the basic structure and trends of the society slowly arose in the 17th and 18th

---

<sup>28</sup> Pearson based his philosophy of science on the German School. Halbwachs was a lecturer in the University of Göttingen.

<sup>29</sup> Arthur Spiethoff is known for his research on the business cycle.

<sup>30</sup> Ferdinand Tonniës became famous for his distinction between *Gemeinschaft* and *Gesellschaft*. The former is organic and instinctive, represented by the

centuries. Censuses were indeed taken centuries ago. The Babylonians and Chinese took censuses in order to levy taxes and recruit armies. The Domesday Book (1086) had also made the assessment of the potential amount of tax as a purpose. William the Conqueror needed a maximal taxation to pay for his vast army. However, the modern regular censuses date back to around 1700 (U.K. Office for National Statistics 2001). In the United States of America, censuses were taken every ten years from 1790. The uniqueness of the USA's census system was its inclusion in the Constitution (Article I section 2 of the 1787 Constitution). Representation and taxes ought to be apportioned among the States (U.S. Bureau of the Census 1989; 2000).

## Context

The censuses were contingent with the societal situation and demands. The history of the census of the United States of America shows the contingent nature and the connectedness of the census with government and society (Rosenthal 2000; Anderson and Fienberg 2001). Three topics are briefly described here; representation, social reform, and immigration. First, the census has always been an instrument to determine representation. The first censuses were used to determine the amount of taxes and representatives for the States<sup>31</sup>. An important issue was the inclusion of the slaves. After a long debate, the *three-fifths rule* was established. Slaves were counted as three fifths of a free man. Secondly, in the second half of the 19<sup>th</sup> century, the census was expanded to include social matters; taxation, churches, pauperism and crime. This extension parallels the increasing social awareness of the society. The data of the census played a central role in the debates about the abolition of slavery. Thirdly, the census was altered to respond to the problems of immigration and migration. Throughout the 19<sup>th</sup> century, numerous waves of immigrants entered the United States of America. The population multiplied by 23 between 1790 and 1910. In particular, the last influx of Catholic, Jewish and Orthodox immigrants from Italy, Poland, Russia and the Balkans was considered problematic. In the 1920s, the introduction of quota was proposed. In the end, the annual immigration for each country of origin was restricted to 3% of the persons counted in the 1910 census as born in the corresponding country. Again, the census data played a crucial role in a major shift in policies - i.e. in population policies. The dependent nature of the census can be found in other countries. British parliament for example only could be persuaded to vote a census act when the burden of a poor population became apparent towards the end of the 18<sup>th</sup> century. The fear of an ever-increasing population crystallized in the work of Malthus (Glass 1973).

---

family or neighbourhood while the latter is deliberative and goal oriented, exemplified by the city or state.

<sup>31</sup> In the 19<sup>th</sup> century, the federal government rarely used its competence to raise direct taxes.

## Example

In 1857, the Southern journalist Hinton Helper published *The Impending Crisis*, an examination of slavery's impact on the South (Helper 1857). The purpose of the study was to influence policies. Helper compares Northern and Southern states in two moments in time -1790 and 1850 and shows that although the northern and southern states started with roughly the same population, the northern states developed much more strongly than the southern ones<sup>32</sup>.

## Relevance

Census taking was surely an important factor in the development of the measurement of government activity. First, like the German historical school, the censuses provided policy makers and scientists with vast quantities of data about society. Meaningful analysis requires meaningful data. Secondly, it was the only instrument to obtain those large-scale data in the era before the development of sampling. In 1827, Quetelet was one of the first to suggest that sampling might be better than censuses. However, he encountered severe and substantial criticism from Baron de Keerbergh (Stigler 2000; p.59). Even today, the replacement of the census by sampling techniques is highly controversial (Stigler 2000; p.4), notwithstanding its proven weaknesses such as the undercount of particular population groups (Gould 1981; Anderson and Fienberg 2001). Finally, census data are widely used in the discourse and the decisions on policies as in the implementation. The abolition of slavery and the migration quota are historical examples. Nowadays, in the United States of America, 200 billion dollars is distributed amongst state and localities based on census data (Petersen 2000). The use of census data in distribution formulas for grants is a global practice.

---

<sup>32</sup> A quote from the report; "In 1790, when the first census was taken, New York contained 340,120 inhabitants; at the same time the population of Virginia was 748,308, being more than twice the number of New York. Just sixty years afterward, as we learn from the census of 1850, New York had a population of 3,097,394; while that of Virginia was only 1,421,661, being less than half the number of New York!. In 1791, the exports of New York amounted to \$2,505,465; the exports of Virginia amounted to \$3,130,865. In 1852, the exports of New York amounted to \$87,484,456; the exports of Virginia, during the same year, amounted to only \$2,724,657.(Helper 1857: pp12-13)"

### 3.10. Social Survey Movement

#### Definition

At the end of the 19<sup>th</sup> century, the vital statistics ran out of steam because of the rise of eugenics and bacteriological discoveries (Désrosières 1998). However, a moral undercurrent has always been present. A moral concern with social conditions fuelled an urge to investigate and change them (Bulmer, Bales and Sklar 1991). This ameliorist reaction against the harsh eugenic theories of society led to innovations in the social research. The social survey was innovative for several reasons. First, it involved fieldwork and the collection of first hand data rather than reliance upon reports by other. Secondly, they attempted to achieve comprehensive coverage within a local area. In contrast, earlier British statistical efforts were of a non-cumulative quality (p.11). Thirdly, the data related to individuals, families and households rather than to aggregates. Herewith, they built on the tradition of the budget studies of Mayhew and LePlay. Finally, the social survey movement, and in particular Charles Booth, provided a breakthrough for social mapping as a way of presenting research results.

#### Context

The social context did not alter significantly compared to context in times of the public health movement. There was still high poverty and mortality rates due to industrialization at the end of the 19<sup>th</sup> century. However, two intellectual currents underpinned the development of the social survey (Bulmer, Bales and Sklar 1991). First, there were the pre-labor 'Fabian Society' and the 'New Liberalism'. These movements expressed a left-wing middle class opinion that more collective solutions to poverty were needed. This opinion was largely inspired by fear. Sporadic social unrest such as the *Trafalgar Square Riots* of 1886 caused anxiety in the middle and upper classes. Secondly, the Settlement movement wanted to do more for the poor than mere relief. They wanted '*to spread education and culture, enable middle class people to form personal relationships with members of the working class, and to discover facts about social problems (Bulmer, Bales and Sklar 1991: p.24)*'. Moreover, the Settlement Movement exported the Social Survey to the United States of America (Platt 1991). The Social Survey Movement was deeply embedded in this reformist agenda.

#### Example

The best-known work of the social survey movement is Charles Booth's study on the '*Life and Labor of the People of London*' (Linsley and Linsley 1993). Booth believed that the poverty debate was underdeveloped because three questions were unanswered; how many people were poor, why were they poor and what should be done to alleviate poverty? Booth determined a poverty line and measured the numbers of people below and above the poverty line. He obtained his data from School

Board visitors who had to enter the homes in their neighborhood to enroll children. Booth based his poverty line on the opinions of these School board visitors. However, he checked the data of the visitors by visiting the neighborhood himself. Other landmark studies were Rowntree's and Bowley's studies of poverty in London that further advanced the methodology (Hennock 1987; 1991). Rowntree no longer relied on secondary sources, i.e. the School Board visitors. He also determined the income of families as a measure for poverty. Bowley for his part used random sampling for his research. Thanks to this technique, the social survey became relatively low-cost. This enabled him to go beyond the local level and make comparisons between cities. In the United States of America social surveys were also conducted.

## Relevance

The relevance of the social survey to the development of the quantification of government is manifold. First, they combined the analytical tradition of the political arithmetic and the public health movement with the pursuit of completeness that was more characteristic for the German tradition. Secondly, they made policy recommendations. The blueprint for the think tanks of the public health movement materialized in the *Settlement Houses* in Britain and the United States of America. Social science was an instrument to realize a social policy agenda. However, the impact on Whitehall in Britain may not be exaggerated (Davidson 1991; p. 360). Although senior researchers of the Social Survey Movement were appointed to positions in the British government, there is little evidence of direct impact on policy-makers. Particularly the *Treasury* and the *Registrars General's Office* were suspicious about the Social Survey Movement and its methods. They saw it as a collectivist device. Thirdly, social mapping proved to be appealing. Graphical reproductions of social conditions had been created earlier. Quetelet for instance made crime maps. However, the Social Survey Movement generalized the concept. The success of Geographical Information Systems in our time proves the enduring salience of the graphical representation of data. Fourthly, the United States of America and Europe were highly connected in the Social Survey Movement. There was intense contact between mainly British and American researchers (Platt 1991). The movement was a direct antecedent of the practices of the New York Bureau of Municipal Research (Williams 2002)

### 3.11. Scientific Management and the Science of Administration

#### Definition

Scientific management and the science of administration brought true innovation to the beginning of the 20<sup>th</sup> century. Although some of the European antecedents, such as the German cameralism, contributed to its development, it was chiefly an American invention (Mosher 1975: p7). Scientific Management and the Science of Administration developed parallel. This is not surprising. The business of running a public and a private administration were considered identical (Mosher 1968). The time studies of Frederic Taylor in the 1890s were a landmark in Scientific Management. He deconstructed a job into elementary actions that could be timed with a stopwatch (Schachter 2004). Scientific management became increasingly popular. Schafritz et al (2004) observes that '*in the first half of this century, scientific management was gospel and Taylor was its prophet.*(p.4)' Meanwhile the science of administration developed along comparable principles. In 1887, Woodrow Wilson provided the object for a science of administration. Administration is the machinery of political life. However, it is raised above the level of mere technical detail by its connectedness with the lasting maxims of political wisdom (Wilson 1887;2004; p.28). The principles of both Scientific Management and the Science of Administration were according to Mosher (1968). (1) *Rationality*: the applicability of the rule of reason. (2) *Planning*: the forward projection of needs and objectives. (3) *Specialization*: of materials, tools and machines, products, workers and organizations. (4) *Quantitative measurement*: applied as far as possible to all elements of operations. (5) "*One best way*": there is one single best method, tool, material and type of worker. (6) *Standards and standardization*: the "one best" ,once discovered, must be made the standard (Mosher 1968; p.72-73). Although the Science of Administration and Scientific Management shared a fundamental view, the concrete practices they developed were sometimes contradictory. Scientific Management for instance supported distributed management, while the Science of Administration advocated a hierarchical executive branch (Williams 2004).

#### Context

At the end of the 19<sup>th</sup> and the beginning of the 20<sup>th</sup> centuries, the rapid industrialization of the United States of America strained the capacity of government that reflected a simpler, more rural way of life (Moe 2003). Regulation was being demanded for industries, transport and urban life. Government institutions therefore needed a professional workforce. Additionally, government was plagued by corruption. The public sector reforms at that time propagated scientific management as a '*war on waste*' (Light ,1997: p18). In response to this challenge, '*the construction of a centralized bureaucratic apparatus was championed as the best way to maintain order during this period of upheaval in economic, social and international affairs. Viewed at this level, the American experience fits a general pattern of institutional development and rationalization in public administration*' (Skowronek in Moe

2003: p11). There was a shift from “*government by the common man*” to “*government by the efficient*” (Mosher 1968). The former refers to a government by the people based on free elections and administration by individuals responsive to the electorate either by frequent election or by immediate dependence upon elected officials (p61). Industrialization and corruption caused a shift towards “*government by the efficient*”. Administration is a profession and a science in its own respect. President Theodore Roosevelt extended the power of the President and the executive branch. This reform also led to increasing deficits of the federal government.

### Example

A landmark study was the work of the “*President’s Commission on Economy and Efficiency (1910-1913)*” which is also known as the *Taft Commission* (Moe 2003). The deficits of the executive branch were chronic and Congress found the ministerial Book of Estimates submitted each year by the Treasury Department inadequate because it did not provide an oversight for the entire government (p.31). The director of the commission, Frederic Cleveland, was also director of the New York Bureau of Municipal Research (infra). The Taft Commission was indeed influenced by the experiences of the Bureau (Kelly and Rivenbark 2003; Williams 2003). The recommendations were mainly agency specific or process based. They were related to micro management. Agencies should be grouped by the purpose they serve and should have uniform procedures. This is in line with the rationale of scientific management. One recommendation had a government wide impact. The Taft Commission proposed to introduce one national budget in order to have a better control on the executive expenditure (Moe 2003). Willoughby (1918), a member of the Taft Commission, described the advantages of the proposed budget reforms. A budget should be an instrument of democracy, for correlating legislative and executive action, and for securing efficiency and economy.

### Relevance

Frank Goodnow’s (1900;2003) distinction between politics and administration fuelled the importance of a separate administrative branch in government. Scientific management and the science of administration was a means to detach administration from politics. Administration and management became for the first time so heavily intertwined. This was also a big step forward for measurement,. Until now, performance of government referred mainly to what we would call outcomes. The indicators chiefly were societal indicators on e.g. poverty, social conditions and crime. The science of administration and scientific management integrated for the first time measurement in the operations of government. They propagated output and efficiency measures. Taylor for instance propagated ‘incentive pay systems’, which is known today as performance related pay (Hood 2000). Although performance measurement was older, performance management largely only came into being in the United States of America at the beginning of the 20<sup>th</sup> century. The first and most influential applications are found in the cities - in particular in the New York Bureau of Municipal Research. The human



relations and later the human resources movement reacted against this search for a one best way of organizing work. Three significant representatives of the movement are Mary Follett<sup>33</sup>, Elton Mayo<sup>34</sup> and Douglas Mc Gregor<sup>35</sup>. The human relations movement, and later the human resources movement put less stress on quantitative performance measurement. Seventy years later, the New Public Management movement, and in particular performance pay initiatives, and resumed the thread of quantification of work.

### 3.12. Cost accounting

#### Definition

A second evolution of that time, besides Scientific Management and the Science of Administration was the development of cost accounting. Cost accounting is in essence the process of tracking, recording and analyzing costs associated with the activity of an organization. Johnson and Kaplan (1987: p126) claimed in their book *'Relevance Lost'* that: "By 1925, American industrial firms had developed virtually every management accounting procedure known today". U. S. business was the world leader from early in the 20th century until about 1970. Modern cost accounting grew to full stature in the early 20<sup>th</sup> century. As ever, there were antecedents. Garner (1968) points to some examples: small workshop owners in 15<sup>th</sup> century Britain, the mines of the Fugger family in Central Europe (1577), the accounts of the Medici in Florence (1431), the accounts the books printed and published by Plantin (1520 - 1589) in Antwerp (Flanders). The industrial revolution in England yielded some innovative practices in the early 19<sup>th</sup> century. The real breakthrough however was in the big conglomerates in the United States of America around 1900.

---

<sup>33</sup> Mary Parker Follett was one of the first authors to stress the importance of human relations for performance of organizations (1924). She distinguishes between power-over and power-with. The former type of power, coercive power is seen as less effective. Real power is co-active power.

<sup>34</sup> Mayo (1933) conducted the Hawthorne experiments. The Hawthorne experiments were set up with the intent of determining the impact of working conditions on productivity (Mayo 1933). The workers improved their efforts for all the conditions, including the control group. This was attributed to the fact that they knew they were being observed. Since then, the term Hawthorne effect has been used to explain change in a situation that occurs merely because an experiment is being run.

<sup>35</sup> Douglas McGregor distinguished between Theory X and Theory Y as two extreme views on motivation. According to Theory X, management assumes employees are inherently lethargic and will avoid work if they can. Theory Y sees employees as ambitious and self-motivated. The satisfaction of doing a good job is a motivation in itself. A Theory Y manager will attempt to

## Context

The 1890s can be called a watershed of American history (Previts and Merino 1979). The change from agrarian simplicity to industrial complexity was complete. Corporations employed over 70% of those working in manufacturing. Technological advances had expanded production possibilities, a nearly complete railroad network permitted mass marketing, and communication improvements facilitated centralization (p.127). A relatively small number of men managed the industrial trusts<sup>36</sup>. Together with industrialization, financial capitalism emerged. Management and ownership no longer had to coincide. A financier like JP Morgan accumulated immense power. The emergence of this type of big business created a new demand for management information for three reasons at three levels. First, the separation of management and ownership triggered demands for the protection of investors. Publicity was seen as a means of external control. Secondly, the vertical integration of mammoth firms holds the risk of bureaucratic inefficiencies, losing the potential gains of the integration<sup>37</sup>. The answer was a centralized unitary structure with high specialization and coordination. Specifically, cost information would meet coordination needs. This is the intra organizational control. Thirdly, the tasks of laborers became increasingly complex. Cost accounting was needed to give managers precise and accurate information about the efficiency of workers engaged in specialized tasks (Johnsen and Kaplan 1987). This is micro control.

## Example

A public sector example of the first cost accounting initiatives was the work of Captain Henry Metcalfe who was responsible for a gun-making workshop. These workshops had a reputation of low productivity (Schachter 1989). Metcalfe published his '*Cost of Manufactures*' in 1885 and reissued it in more successfully in 1890. He prescribed that a card should be made out for almost every possible type of transaction or transfer of material (Garner 1968). Each card, moreover, had a space for pricing the article used, as well as the order number to which it was to be charged. The prices of orders or of component parts could be shown on every recording. His study laid the groundwork for both the development of cost accounting and the implementation of many operational systems. In the private sector, the *Du Pont de Nemours Powder Company* (founded in 1904) developed a highly centralized

---

remove the barriers that prevent workers from actualizing their potential (McGregor 1960).

<sup>36</sup> Some leading industrials were William Vanderbilt (1856-1938) - railways, John D Rockefeller (1838-1937) - oil, John Pierpont Morgan (1837-1913) - financier and banker, Andrew Carnegie (1835-1919)- steel, the Du Pont de Nemours family - chemicals, William Durant (1861-1947) Alfred Sloan (1875-1966) and Henry Ford (1863-1947) - automobiles

<sup>37</sup> Horizontal integration; both firms produce the same commodity or service. Vertical: a firm acquires either a supplier or a customer.

and integrated information system. The leading measure was the return on investment (Johnsen and Kaplan 1987).

## Relevance

The early development of cost accounting was remarkable for several reasons. First, it was a joint venture of the public and the private sector. The work of Henry Metcalfe and the New York Bureau of Municipal Research -both public sector- was seminal, as well as some applications in the large companies of that time. This is not surprising. Claims of control and openness echoed in the public sector and in the private sector (Previts and Merino 1979). Cost accounting institutionalized in the private sector. In the public sector, it is still considered innovative in most OECD countries (Pollitt and Bouckaert 2004). Secondly, cost accounting was closely linked with scientific management and they mutually reinforced each other. Taylor and Metcalfe actively debated their concepts and techniques at the 1886 meeting of the *American Society of Mechanical Engineers*. Their writings were very influential in both the public and the private sector at the time (Schachter 1989). Thirdly, on the technical level, performance measures are integrated in the financial system through cost accounting. The financial system could be seen as the most institutionalized information system in modern organizations.

### **3.13. New York Bureau for Municipal Research<sup>38</sup>**

#### Definition

Taylorism can be found outside the factory too (Schachter 2004). One of the most prominent developments was the effort of the New York Bureau of Municipal Research (Schachter 1989; Williams 2003; Williams 2004). Together with cost accounting and the scientific management/science of administration, it was the third distinct innovative movement around 1900. The three co-directors of the Bureau were William H. Allen, Henry Bruere and Frederick Cleveland. The New York Bureau of Municipal Research developed many of the performance measurement concepts that are in use today. Data collection was embedded in accounting practices. They propagated the development of cost accounting. Record keeping efforts were directed towards work records such as time sheets and work plans as well as output and outcome indicators. These indicators were supplemented by studies of social indicators and needs. The information was to be used for several purposes. First, the information was used for reporting. Reporting of efficiency and effectiveness, unit costs and gains and losses was an important device for making the operations of the cities transparent for citizens.

---

<sup>38</sup> This section is based on the extensive research by Daniel W. Williams (2002, 2003, 2004).

Secondly, the information was integrated in the budget. A functional budget was developed in order to compare similar work units. Thirdly, the information was used for productivity improvements. This drastically reduced waste.

## Context

The general societal context is described in the previous section on the Science of Administration and Scientific Management. Rapid industrialization and urbanization put pressure on a government that was plagued by corruption and waste. One of the main changes was the strengthening of the elected executive. This section focuses in particular on the reason why the first and most far-reaching initiatives were taken in the big cities on the East Coast and worked their way up to the state and federal level in the next decade. Virtually all European administrative thought, literature and education on the contrary began with the nation-state and worked downward (Stone and Stone 1975: p.8). The structure of the American government explains why administrative reform first flourished in the cities. The cities were at the time the biggest governments. They had to provide water, fire protection, remove waste, protect health, control lawlessness, pave streets and provide education (Stone and Stone 1975: p.17). However, the cities were under the control of political cliques and bosses. Most positions were filled based on political patronage. In many cases, entire families were dependent on the bosses. Moreover, they accepted bribes, sold franchises and constructions of public facilities in areas where the political bosses bought land, and accepted kickbacks in the award of contracts. In short, the whole spectrum of political corruption could be found in municipal government at the turn of the 19<sup>th</sup> century.

## Example

An exemplary study is William Allen's research on hospital efficiency (Allen 1906). It is a good illustration of the integration of cost accounting and scientific management in a public sector setting. The goals are inherently public and go beyond profit making. A large part of the text is devoted to the persuasion of critics of his approach. This was surely a necessity. Allen observes that '*many managers looked askance at the proposed remedy -viz.: uniform, up-to-date system of accounts and reports. "What! Remove a deficit by expending more on statistics?"*' (Allen 1906: p.300). He extensively discusses the shortcomings of several hospitals. A short excerpt is replicated here. '*Many of them failed to give even the total patient beds, (...), several of the important hospitals did not give even the number of patients; (...); eleven of the fifty-one gave the gross cost per patient per day, one the cost of food per patient per day; five, the number of days' board given employees; two, the number of private patients*' (Allen 1906: p.306). Finally, Allen discusses of a new uniform framework agreed upon by the four leading hospitals in New York. It consists of 120 items. Hospitals had to make a statement of all their expenses on the items. Five out of six schedules propose a functional classification of the budget. The first schedule details the expenses. Items are for instance the expenses for salaries and wages of

physicians, nurses, wards, etc.; the equipment for nurses such as books, uniforms and instruments, and so on. The second schedule represents the revenues. The third, fourth and fifth schedule number the financial transactions, provide a comparative balance sheet and give a statement on the capital funds. The sixth section provides comparative statistics on patients admitted, discharged and costs per patient. The classification is a modern accounting system as well as a performance measurement system. Allen was optimistic about this new instrument. He believes that with this tool *'it will not be true that eighty hospitals spend \$4,000,000 without the community's learning anything from their experience to make tomorrow better than today (Allen 1906: p. 311)*.

## Relevance

The work of the New York Bureau of Municipal Research was a landmark in the development of modern performance measurement. For the first time, the quantification of government was about the administration (some would say the machinery) of government and not only about the social conditions on which government was expected to have an impact. This practice was widely disseminated and therefore served as a leverage for reform in government in the United States of America (Williams 2003). The dispersion of the ideas of the bureau occurred along four tracks. First, there was a strong connection with the academia. The bureau's approach was integrated in the curricula of political sciences albeit gradually<sup>39</sup>. Secondly, the bureau intentionally exported its work through service to other communities and active promotion. Many more bureaus were set up throughout the country. Thirdly, the bureau sought active contact with officers and administrators. Fourth, the bureau promoted its ideas in the federal government. The Taft commission was the first major effort of that kind.

### **3.14. Performance Budgeting, PPBS and their successors (MBO,ZBB, GPRA)**

#### Definition

The first Hoover commission was a milestone in public administration (Moe 2003; Shafritz et al. 2004). The fifth finding of the first Hoover Commission stated *"the budgetary processes of the Government need improvement, in order to express the objectives of the Government in terms of the work to be done rather than in mere classification of expenditure (The Hoover Commission report in Shafritz et al 2004: p.162)"*. This was the mission statement of performance budgeting. Performance budgeting

---

<sup>39</sup> The New York Bureau of Municipal Research founded the Training School for Public Service. The universities responded later. The University of Michigan set up a one-year master in 1914. Columbia University promoted the research of the Bureau as early as 1915, but did not formalise the program. Stanford (1921), Syracuse (1924), Southern California (1928), and the

however expanded a great deal with the introduction of the *Planning Programming Budgeting Systems* (PPBS). New program expenditures had to be weighed against the marginal benefits of each program. Schick argues that the addition of the management function to the control and management function is the main novelty of PPBS (Schick 1966). He also presents it as '*a radical change in the central function of budgeting, but anchored to half a century of tradition and evolution (p.243)*'. PPBS may indeed be seen as the last step of the New York Bureau of Municipal Research's journey in American public administration. PPBS was a success in the Ministry of Defense. It is still in use in the military nowadays. In 2003, the system was reformed including a biannual budget process. Now it is called the Planning Programming Budgeting and Execution System (McAffery and Jones 2004). However, the transfer to other departments was problematic. Amongst others Aaron Wildavsky (1969) forcefully attacked the system. He demonstrated that the planning and analytical functions of PPBS were contradictory to the essential nature of budgeting (Shafritz et al. 2004). The Office of Management and the Budget officially threw in the towel in 1971 (Kelly and Rivenbark 2003). Performance budgeting did not reappear until the Government Performance and Results Act (1993). A final note on Europe. PPBS was widespread in Europe. Great Britain introduced it in the Ministry of Defense in the late 1960s and then began to extend it to other departments, particularly in education and science. In France, the government decided to apply the system in 1968, first in the Ministry of Defense and then in relation to energy, town planning, and departments such as posts and telegraph. By the early 1970s, PPBS had become an integral tool of national economic planning.

## Context

The context may well explain the development and the nature of performance budgeting. First, the development of performance budgeting was a response to the fragmentation of the executive branch in the early years of the Great Depression (Moe 2003; Kelly and Rivenbark 2003). After the stock market crash of 1929, unemployment soared and the whole banking system was at its last gasp. The confidence in a self-regulating economy vanished rapidly. Franklin Delano Roosevelt's administration responded to the crisis by creating many programs and agencies in a for that time unorthodox manner (Mosher in Kelly and Rivenbark 2003: p 28). New agencies were established for new programs independent of existing departments and agencies in the same field. Personnel was recruited without reference to the established civil service. Funds were appropriated outside the regular budget process leading to a "double budget" system. The division between politics and administration was challenged since most initiatives came from the administration. Mosher notes, "*for orthodox students of public administration, the first new deal was chaos (p.28)*" The Brownlow and Hoover committees sought to re-establish executive control through a clear line of executive authority. Performance budgeting was one of the instruments to realign executive control and managerial freedom. Secondly, system thinking

---

Brookings Institution (1927) established formal public administration programs (Stone and Stone 1975).

influenced the nature of performance budgeting. In fact, PPBS is seen as a tool for resource allocation along the lines of system theory (Shafritz et al. 2004;p190 ). For the field of Public Administration, and Kahn's book the social psychology of organizations (1966) was particularly influential. They applied the open systems theory of Ludwig von Bertalanffy to organizations. They argued that organizations are open systems that get their energy from the environment. If organizations are to survive, they have to be open and responsive to their environment. This is especially the case for environments that are changing rapidly. The Planning Programming Budgeting System should enable organizations to build this capacity for responsiveness.

### Example

Churchman and Schainblatt (1969) describe a PPBS example within the Californian State Department of Public Health. PPBS is applied to the mission of combating alcoholism. Six programs are identified including the prevention of alcoholism, the restoration of early-stage alcoholics, the care of chronic alcoholics, etcetera. Sub-programs are formulated for each program. The program 'restoration of early-stage alcoholics' for instance has five sub-programs including the 'diagnosis evaluation and referral' and the 'medical treatment' of alcoholics. The sub-programs for their part are broken down to display the different ways in which they can be carried out. The medical treatment may be accomplished either in specialized emergency-care centers or in general hospitals. After the formulation of the program structure, a cost-benefit analysis is applied to estimate the relative merit of the programs. The interrelations of different program elements too should be identified. In order to do that, the financial contribution of different organizations to the programs needs to be mapped. For instance, what is the contribution of the department of Education and the Department of Employment to the prevention of alcoholism? Additionally, the size and the different resources for the different target groups may be mapped along the program structure. A typical question for the detection of alcoholism is the size of the target group of non-alcoholics under 21 and above 21 and the resources needed to get to the target group. Finally, a multi-year program and financial plan for the program-structure is proposed. In the case of the anti-alcohol program of California, the time span was five years.

### Relevance

Performance budgeting and its sister PPBS have been heavily criticized. Wildavsky (1969) argues persuasively that in particular the fixation on the program structure is pernicious. There is not sufficient analytical capacity to provide a meaningful program structure for all the activities. Moreover, causality between programs is not explored sufficiently. The weighing of programs in that way becomes difficult. He quotes William Gorham to point to the conflict between analysis by analysts and the value judgments of politicians. The former cannot resolve the problems of the latter. With his criticism, Wildavsky does not only attack PPBS, but also the holistic system approach to public administration.

Despite the criticism, PPBS was in hindsight a significant movement for the development of performance measurement. First, the output and effect orientation of PPBS was for some years actively pursued in governments in the United States of America. Even though the system component was a victim of the over-commitment (Hood 2000), many bits and parts lingered (U.S. General Accounting Office 1997). Not only policy analysis was rescued from PPB, performance budgeting too persisted. Subsequent initiatives such as Management By Objectives (MBO), Zero Based Budgeting (ZBB) and the Government Performance and Results Act (GPRA) prove the continuous salience of the concept (Kelly and Rivenbark 2003). Secondly, performance budgeting and PPBS were the first performance-oriented reforms that have been extensively exported internationally. Still hypothetical, performance budgeting might have primarily been exported through the Marshall plan to post-war Europe.<sup>40</sup> In any case, PPBS practices were found in Australia, Austria, Belgium, Canada, France, Ireland, Japan and the United Kingdom (Novick 1973). The export of PPBS continues today in the context of the enlargement of the NATO. All the candidate members of the NATO have at least formally adopted PPBS for their armed forces (Geneva Centre for the Democratic Control of Armed Forces and the Ministry of Defence of Bulgaria 2001).

### 3.15. Social Indicators and Quality of Life

#### Definition

The late 1960s and early 1970s the scientific community (both academic institutions and research bureaus in and around government) developed an interest in social indicators. *'The attempt was made to construct standard measures of the state of health, crime, well-being, education and many other social characteristics of a population (Bulmer 2001: p468)'*. The definition of a social indicator has four dimensions (Cazes 1974). First, it is a measurement of a social phenomenon (and thus is susceptible to all the potential measurement errors). Secondly, it is trans-economic. Social indicators aspire to assess the social gains and losses that escape market calculations. Thirdly, a social indicator has a normative character. It quantifies social goals. Social indicators should enable the assessment and benchmarking of progress towards the values and goals (Bauer 1966). Fourthly, the social indicators are an integrated system. They should provide *'comprehensive and balanced judgments about the conditions of major aspects of a society'* (Mancur Olson Jr. cited in Cazes 1974; p.20). Like with PPBS, system thinking affected the early works on social indicators. Although there were some antecedents<sup>41</sup>, the breakthrough of the social indicator movement can be traced back to a study of

---

<sup>40</sup> Daniel Williams formulated this hypothesis during the meeting of the Study Group on Quality and Productivity in the Public Sector of the European Group of Public Administration in Ljubljana, September 2004.

<sup>41</sup> Particularly noteworthy is the President's Committee on Recent Social Trends (1929) installed by President Hoover. The final report ran to 1700



side effects of the NASA space program. However, the data availability to assess the second order consequences of the program was largely insufficient. This was the conclusion of a book on the study of the NASA program edited by Raymond A. Bauer (1966). The title of the book -'social indicators' - became a familiar catchword and branded the movement. After its inauguration in the 1960s, many governments in different countries and from different tiers took part during the 1970s<sup>42</sup>. Some international institutions such as the United Nations and the OECD joined the cause (United Nations 1975; OECD 1982). Although the social indicators movement stagnated during the 1980s, the 1990s showed a revival (Noll and Zapf 1994; Noll 1996).

## Context

The rise and rapid diffusion of the social indicators was fostered by the societal climate of the late 1960s and early 1970s (Innes de Neuville 1975; Thiry 1980; Noll 1996; Dowrick and Quiggin 1998). After almost two decades of economic growth and prosperity, for the first time, there were questions about the limits of economic growth. What are the social costs of economic progress? Is quantity pursued at the expense of quality? President Lyndon Johnson stated in 1964 that '*the great society is a place where men are more concerned with the quality of their goods than with the quantity of their goods*'. The social indicators movement was largely a response to the perceived preponderance of the economic indicators such as the GDP - termed "*economic philistinism*" (Bauer 1966; ix). Secondly, particularly in continental Europe, the welfare state was developed in the decades after WWII. The welfare state is a concept of government in which the state plays a key role in the protection and promotion of the economic and social well-being of its citizens. It supports an active social policy. The development of the welfare state triggered the demand for social data. The economic crises of the second half of the 1970s and the cutback management of the 1980s may explain why the movement ran out of steam during that era (Bulmer 2001).

## Example

A typical study is the assessment of the Quality of Life in US metropolitan areas commissioned by the Environmental Protection Agency (Ben-Chieh and Livingston 1975). The study presents a set of indicators that mirrors the social system of respectively the large, medium and small metropolitan areas in the United States of America. The quality of life has five components. First, there is an economic component, which consists of the personal income per capita, and the community economic

---

pages, 29 chapters and 2 volumes. It used statistics to describe how America had changed between 1890 and 1930.

<sup>42</sup> Some examples include Austria, Belgium, Denmark, Federal Republic of Germany, France, the United Kingdom, the Netherlands, Norway, Sweden (Thierry et al. 1977; Thiry 1980; Noll 1996)

health. The traditional economic indicators are supplemented with measures of income equality, poverty, unemployment, etcetera. Secondly, the political component deals with the connectedness to the media, voter turnout, salaries of teachers and firefighters, crime and health and expenditure on public welfare. The political dimension is defined rather freely. Thirdly, the environmental dimension incorporates indicators on air, visual noise and water pollution as well as climatological data and data on recreation areas. Fourthly, the educational and health component includes infant mortality, death rate, the provision and accessibility of medical care, and the level of schooling. Finally, the social component is a rest category with for instance data on mobility and car ownership (the latter at the time being a positive factor), population density, participation of colored people and women in the labor market. The scores on the different components, based on the group means, are calculated and the cities are ranked from outstanding to substandard. Some illustrative conclusions are the following. First, the quality of life is in general lower in the southern states. However, there are many good metropolitan areas in the south as well as there are substandard ones in the north and the west. A policy therefore should be designed for individual areas rather than groupings of metropolitan areas. Secondly, economic welfare does not necessarily imply an improved quality of life. Economic growth is not enough. Thirdly, health and education scores differ seriously, which leads the author to raise concerns about the uneven development of Human Resources in different metropolitan areas. The study is illustrative for the social indicator research because of its system approach, its normative dimension, and the link of policy advice to the findings.

## Relevance

The social indicator movement has failed in its ambitious aims (Andrews 1989; Cobb and Rixfors 1998; Bulmer 2001). The failure was partly caused by the shifting politico-economic context in the 1980s. However, there were intellectual explanations too. The most important reason was the lack of a common unit of measurement to determine value that might have compared to the measure for value in economic measurement (i.e. money). The movement failed to standardize its work - both intra and cross-national. An example of the latter is the bifurcation of the conceptualizations of welfare by Scandinavians on the one side, and Americans and Western Europeans on the other side. The former stressed the objectively measurable living conditions while the latter looked at the subjective perceptions of welfare (Noll 1996; Bulmer 2001). The sought unified system of social accounting too did not materialize. Still, the social indicator movement is undoubtedly relevant for the development of performance measurement. First, the statistical apparatus of governments was expanded to cover more phenomena. New time series have been developed. Secondly, the extended statistics on the social condition of the population allowed performance measurement systems to cover better the outcomes of government action (Aristigueta 2004). Prominent authors stress the importance of outcomes (and not only output) in the public sector (e.g. Hatry 1999). The outcomes need to be framed in a logic model of the field of action of the organization. The social indicators provide an important addition either as an outcome indicator or as an independent (contextual) variable explaining a dependent outcome. Thirdly, notwithstanding the criticism, the idea of compiling sets of

indicators to assess the quality of life keeps its salience (Aristigueta 2000; Aristigueta, Cooksy and Nelson 2001). In the 1990s, quality of life indicators revived in particular in local government, often under the banner of 'sustainable community development' (Innes and Booher 2000; Hill and Wegener 2002).

### **3.16. The Quality Gurus**

#### Definition

While the social indicator movement aspired to substantiate the outcome side of performance measurement, the quality movement was aiming at the management side, i.e. the input, processes and output. Joseph J Juran (1951), W. Edwards Deming (1986), Philip B. Crosby (1979) and Armand V. Feigenbaum (1956) were the apostles of the movement (Bowles and Hammond 2001). The quality models were developed in the 1950s and the 1960s. In the 1970s and 1980s, they were implemented in the Japanese industry on a large scale (Bouckaert and Thijs 2002). The success of the Japanese economy lent the quality movement its credibility. In the 1980s, it was imported in the United States of America; first in the private sector and later in the public sector. Quality management is allegedly a new style of management (Amsden et al. 2001). Six attributes distinguish it from conventional (read Taylor, Fayol and Barnard) management. First, quality management focuses on customers rather than on profits and efficiency. Secondly, quality management is a holistic system that goes beyond a single management function. Thirdly, the organization should strive towards continuous improvement, rather than establishing the one best method for all time. Fourthly, organizations that apply quality management empower people. Responsibilities and opportunities are not only a matter of the top of the organization. Fifthly, all the members of the organization have to take part in the planning and control cycle. Hierarchy is not the best way of controlling. Finally, senior management has to create a facilitating environment that fosters the above five attributes through dynamic leadership. Since organizations are open systems, senior managers have to attract the external resources to keep the organization thriving.

#### Context

Two contexts at two points in time are important for the quality movement; one at its conception in the 1950s and 1960s and one at its breakthrough, in the 1980s. First, when the quality models and principles were developed, system thinking was the dominant intellectual framework. Features of the systems theories such as the management of interdependent relationships and the characterization of the environment are consistent with the quality movement (Ehrenberg and Stupak 2001). Secondly, after an incubation period in Japan, the quality movement finally broke through in the United States of America and Western Europe in the 1980s. At that time, the economic crisis of the late 1970s and the 1880s necessitated a cutback management (Kelly and Rivenbark 2003). The level of taxation was

believed to have reached its ceiling. Public administrators found themselves working in conditions of fiscal stress in which they tried to accomplish unlimited goals with fewer real resources (Caiden 1981). President Reagan stated that government was the problem rather than the solution, which implied that less government would lead to fewer problems. The Grace Commission, officially *the President's Private Sector Survey on Cost Control* (1982-1984), put forward a "War on Waste" (Moe 2003). Bouckaert (1990) labels the 1980s as *Government by the Private Sector*. Meanwhile, both business and government leaders looked at the success of the Japanese economy. NBC broadcasted in 1980 a documentary about Deming titled "If Japan Can...Why Can't We?" This documentary was the starting shot of the quality movement in the United States of America. First, it was adopted in the private sector, later it was translated to the public sector. In the 1980s, this was not considered a big step, since public and private was seen by many as fundamentally alike in both unimportant and important respects. Mimetic pressure from Japanese practices and the private sector advanced the implementation of private sector principles and practices in government.

### Example

Cases of quality management implementation are abundant and picking an example is a bit arbitrary. Here we look at the case of the Inland Revenue System (IRS) in the United States of America, since the practice is accurate and concisely documented (Mani 2001). The immediate cause for the IRS to adopt quality management was a backlog in the processing of tax forms in 1985. First, senior staff was trained in quality planning, improvement and control. In addition, an explicit commitment from the executives was sought. In 1987, the participation expanded throughout the organization with the assignment of quality coordinators and quality teams, extended training and the stimulation of a customer-orientation. The effort was continued in the 1990s. The results of the effort are unclear. There was no significant rise or fall in productivity. However, Mani (2001: p.673) finds that quality management made a difference in the IRS's culture. Moreover, customer-satisfaction too improved. There is more employee involvement and quantitative analysis used more generally.

### Relevance

As with most movements, the quality movement was also criticized (Harari 2001). The critics reproach the quality movement to create a new bureaucracy. That it is a matter of whiz kids in the ivory tower that is called the quality department. It is called faddism, egotism and quick-fixism (p.748). Critics address the transfer of models from private to public (Swiss 1992). The customer in the public sector is difficult to identify. Output in the public sector does not straightforwardly lead to outcome. The government culture is problematic for quality management, since the incentive structure and the constraints for top managers are not fostering process improvement (Wilson 1998). However, the quality movement had and still has an impact in the public sector. Public managers still like to use quality models. Three issues may be raised here. First, quality models involve performance

measurement. Quality models propagate the measurement of processes and outputs. In many organizations, they steer measurement efforts in that they point managers to aspects of the organization that might need measurement. Quality models provide a helicopter view of the organization and are a steppingstone for measurement. Secondly, the models were altered and made more apt for the public sector. Two popular quality models, the Balanced Scorecard and the European EFQM<sup>43</sup> model progressively included attention for outcomes. Some years ago, the member states of the European Union launched the Common Assessment Framework, a quality model that is distinctively designed for the Public Sector. Thirdly, the quality models spread globally. They were promoted by international consulting firms and mostly were connected with ICT reforms. Many computerization projects are coupled to models such as a Balanced Scorecard.

### 3.17. New Public Management (NPM)

#### Definition

The next set of management reforms, unified under the flag of the New Public Management, spread around the Western world in the 1990s. The movement is still influential at present and this chapter could be written in either the past or present tense. Osborne and Gaebler's (1992) *Reinventing Government* and the *National Performance Review* (Government Printing Office 1993) were two main texts that launched the NPM in the United States of America. The Public Management (PUMA) section of the OECD heavily promoted the application of the NPM in its member states. The roots of the NPM however lie in Australia, New Zealand and the United Kingdom (Zifcak 1994; Flynn and Strehl 1996; Schafritz et al. 2004). NPM doctrine prescribed that public agencies should be split up in small policy oversight boards and larger performance-based managed organizations for service delivery. The latter organizations should compete with private sector organizations. *Performance* should be the criterion to evaluate agencies and thus, performance needed measurement in an all-inclusive way. Under the colors of the public management resides a broad array of management tools, of which the compatibility is often contested (Hood 1991; Williams 2000).

#### Context

The intellectual foundations of NPM are Public Choice theories on the one hand and businesslike '*managerialism*' on the other hand (Hood 1991; Ferlie et al. 1996). Public Choice Theory is directed toward the study of politics based on economic principles. People are assumed well-informed maximizers with logically consistent preferences and an instrumental and self-regarding behavior (Dunleavy 1991; p.3). Public Choice explains phenomena with the individual preferences of voters and

---

<sup>43</sup> EFQM: European Foundation for Quality Management

bureaucrats (Buchanan and Tullock 1962; Niskanen 1971). Managerialist approaches prescribe that if government would be run like the private sector, it would be more efficient. The new practices reflect these intellectual frameworks (or vice versa? – see Brunson 2002; p169<sup>44</sup>). The objectives of the NPM instruments are mainly directed towards increased accountability of the administration (not politics) on the one hand and increased efficiency on the other hand. A better accountability should avoid that people would vote with their feet (Hirschmann 1970). The pursuit of more efficiency was consistent with the managerialist approach. The societal context favored both the intellectual as the practical developments. Two trends stand out (Zifcak 1994; Pollitt and Bouckaert 2004). First, the fiscal crisis of the eighties forced governments throughout the world to save on their public sectors. The motto was ‘to do more with less’. Secondly, the decline in trust of citizens in government pushed governments to be more responsive. Service charters, ombudspersons and citizen-participation have been set up. Practical applications reflected the societal and the intellectual context (Figure 7). The specific socio-economic conditions made the balance tip towards the accountability or the efficiency side.

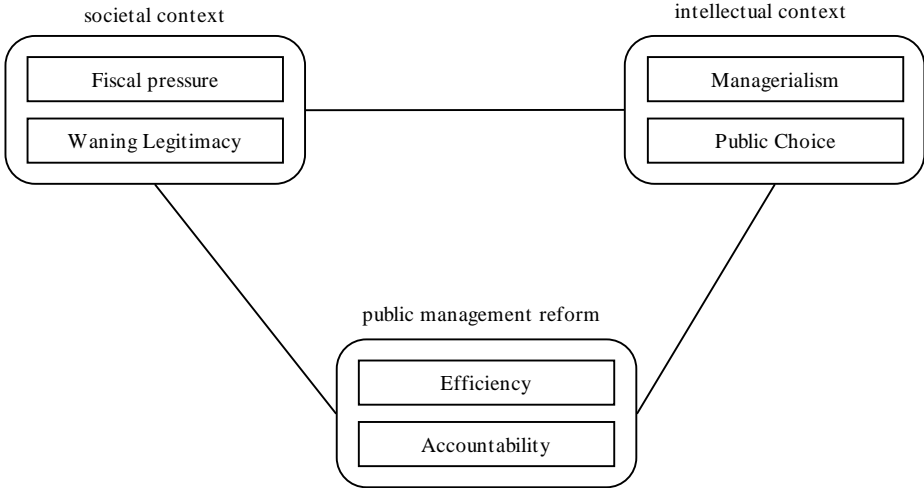


Figure 7: The NPM triangle: reforms, societal context and intellectual frameworks

Example

There are numerous examples of NPM reports by governments. As an example, we look at New Zealand, the Mecca of the NPM (some would say Eldorado). The reforms were far reaching (Schick 1996; Pollitt and Bouckaert 2004). The eight major changes to the system are summed up (Schick 1996: p10-11). First, financial statements, the budget, and appropriations are on an accrual basis.

---

<sup>44</sup> Brunson distinguishes between talk (ideas, plans, ideologies) and action (implementation). He sees both as activities with a value in its own right. The two activities may be decoupled, ideas may control action (the

Commercial accounting standards are applied to all public entities. Secondly, departments prepare monthly financial reports, quarterly performance reports on their purchase agreements, half yearly reports on the chief executive's performance agreement, and an annual report on financial results and outputs. The annual report is audited. In addition to departmental reports, the government issues a combined financial statement. Thirdly, appropriations for operating expenses are made by output classes. The latter classification is not a supplementary schedule but the main form of appropriation. Fourthly, chief executives are appointed under term contracts. Public employees work under individual or collective employment contracts. Fifthly, managerial discretion is less constrained. Within budget limits and law managers are free to select the mix of inputs to be used in producing agreed outputs. Sixthly, accountability for resources and results is maintained through contract-like arrangements within government (performance and purchase agreements). Seventhly, a capital charge is levied on the value of each department's physical and financial assets, net of liabilities. Eighthly, departments maintain their own bank accounts and are responsible for managing cash balances. They earn interest on these accounts. New Zealand's NPM reforms were the most comprehensive. They are considered to be the archetype of the NPM movement.

## Relevance

Nowadays, at least in the academia, the post-NPM era has dawned (Minogue, Polidano, and Hulme 1998; Box et al 2003, Kickert 2003). It may be too early to assess the relevance of the movement for public administration. Results can only be assessed through a glass darkly (Pollitt and Bouckaert 2004). The potential relevance for the quantification of government lies mainly in the attempt to integrate measurement in public management. The use of performance information was not constricted to policy advise or budget and planning documents. It was integrated in many management functions such as human resources management, facility management, public private partnerships, privatization and the controlling of agencies. This contractualism and devolution did have an impact on political oversight (Christensen and Laegreid 2001). Performance measurement was pivotal to many reforms and many measurement initiatives were taken by administrations. However, the difficulties with NPM risk contaminating the pursuit of quantitative information about government at work. It remains to be seen whether the overall impact of NPM on performance measurement will be positive.

---

intellectual foundations precede action), or ideas may explain action (the intellectual explanation is given in hindsight).

### 3.18. Evidence Based Policies

#### Definition

A last movement we would like to discuss in more detail is the Evidence Based Policy Movement (EBP). The EBP movement until now has predominantly been of a British origin (Solesbury 2001) and was initially pursued in the medical and public health sector (Davies and Nutley 2000; Black 2001). By the end of the 1990s, EBP has spread in virtually all policy sectors. Moreover, gradually, the terminology is finding international acceptance amongst others in the European Commission's White Paper on Governance (European Commission 2001a). The Cabinet Office (1999) defines the mission of EBP as follows: *"government should regard policy as a continuous, learning process, not as a series of one-off initiatives. We will improve our use of evidence and research so that we understand better the problems we are trying to address. We will make more use of pilot schemes to encourage innovations and test whether they work (Cabinet Office, 1999: p.17)"*. Sanderson (2002) distinguishes two main forms of evidence. First, there is evidence to promote accountability in terms of results. Evidence in this case comes in the form of information. It has led to an increased use of performance management in government with league tables and targets. A second type of evidence is to promote improvement. Here, evidence is what Sanderson (2002) calls 'knowledge' (p.3) that is acquired through research.

#### Context

Solesbury (2001) identifies three conditions that furthered EBP movement in the UK. First, there has been a utilitarian turn in research funding policy and practice. Research should not only lead to understanding of society, it should also offer guidance on how to make things better. Solesbury (2001) identifies this policy in both the research charities and the Economic and Social Research Council. An additional pressure for researchers to make their research readily usable is the competition from commercial research and consultancy. Secondly, Solesbury (2001) observes a decline in confidence in the professions. He speaks of a 'retreat from priesthood (p.6)' of professionals. This decline in public trust would instigate professional associations to seek support for their practice by evidence. Thirdly, the coming into office of New Labour supported the EBP movement. The white paper on modernizing government (Cabinet Office 1999) championed a replacement of ideology by pragmatism. Policy was said to be founded on evidence about what works rather than ideological predispositions.

#### Example

The Sure Start – program is supported as one of the prime examples in evidence based policymaking (Glass 1999; Eisenstadt 2000). The aim of the program was to ensure that children living in poverty



have a better starting position in school. Sure Start was area based and the districts were chosen based on indicators of deprivation. The planning was based on nationally set objectives and targets with a mix of outcome and input indicators. Research evidence was a key to developing the arguments for 'investing in very young children, deciding on the key principles and determining the activities that the program funds at the local level (Eisenstadt 2000: p.7). An example of the latter is the decision to mainly support initiatives that worked with both parents and children, because they were proven more effective. The informative component of the EBP movement with indicators and targets, and the knowledge component of the EBP movement with the use of social research, as identified by Sanderson (2002) (supra), are showing in the Sure Start program.

## Relevance

The evaluation of EBP is taking place at the present. The critics of the EBP movement mainly attack some allegedly naive and unduly rational assumptions. Young et al. (2002: p.15) call it new instrumentalism. They fear that research will only be responding to policy demands. In this way, the enlightenment function or research will be lost. Marmot (2004) makes a similar argument when he asserts that action precedes research rather than the other way around. He wonders whether we should talk of policy-based evidence rather than evidence based policy. Sanderson (2002) points to the modernist rational angle of the EBP movement. This objection is also reflected in practice when educators claim that education is more complex and culturally specific compared to medicine and health care – the sectors where EBP originated. Therefore, causation is said to be more difficult in education (Davies 1999). The potential of the EBP movement for performance measurement mainly lies in the development of policy indicators that complement the managerial focus of the NPM based management systems. It is still too early to assess whether this potential has materialized.

## 3.19. Conclusions

In the Anglophone world too, the French phrase *plus ça change, plus c'est la même chose* is in vogue. It suggests that most change is superficial spin while the bottom line remains untouched. Mintzberg (1993) for instance showed that it is always our own age that is turbulent and that therefore turbulence is normalcy. Does this apply to measurement in and of the public sector too? Are recent measurement efforts normalcy rather than change? We tend to conclude that some genuine changes can be tracked down throughout history. The most notable evolution is the increasing integration of measurement in the core of the public sector. The quantification of government started outside government. It was usually the third sector that began to measure results of the public sector as a tool for advocacy. External actors inspired early movements such as the political arithmetic, the public health movement, and the social surveys. The German statistics are a noteworthy exception. The purpose of these measurement efforts was to provide the ruler with a '*mirror of the state*'. The 20<sup>th</sup> century witnessed an increasing integration of measurement within and by the public sector itself. Quantitative approaches

to policy and management -in this order- became an inclusive part of government. First, the elaboration of statistical systems led to government-wide existence of policy indicators. The social indicator movement in particular promoted comprehensive sets of indicators for all aspects of public action. Secondly, after the development by the *New York Bureau of Municipal Research* and an incubation period of nearly a century, NPM was the first inclusion of quantitative information in public management on a government-wide and international scale.

Why is almost every alleged revolution or paradigm change written off as a fad after several years? A recurring phenomenon is the over-commitment to newness. In the early days, the reforms need to be sold (Sieber, 1981). In their sales pitch, change agents usually promise more than what reasonably can be expected. Nathan (2004) argues that the whole alphabet soup, PPBS, MBO, ZBB, NPR, and GPRA, is victim of the inclination to overselling. Other movements such as the quality movement and the social indicators faced the same criticism (Bulmer 2001; Harari 2001). In most cases, the reforms are a part of dialectic process whereas the reform is a strong antithesis against an existing situation (thesis). The reaction against the reform presumably is part of the formulation of a synthesis. The social indicators for instance were a reaction against the so-called '*philistinism*' of the economic indicators. The synthesis was a more balanced statistical system that included some (not all) of the social indicators. Good expectations management of the reform may limit the disturbance that reforms bring along. A synthesis may be attained with fewer reform costs. The latter argument makes the case for historical research. Organizational memory loss may be avoided (Pollitt 2000). Although it is not feasible or desirable to remember everything, the consequences of highly forgetful organizations may be an intermittent and poor policy and management learning -re-trying solutions which failed 5 or 10 years before.

New movements usually present themselves as an antithesis against mainstream practices. Nonetheless, movements that take hold do not come out of the blue either. They are embedded in a societal context. Generally, they express societal problems and provide answers. The answers typically appear straightforward (simplistic, some would argue), not bothered by practical implementation. The description of the context in which movements arise, demonstrates the importance of the fit of a movement with its environment. The breakthrough of TQM in the public sector in the 1980s for instance fitted the problems of that time. Government was bankrupt and the private sector, in particular in Japan, was the guiding light. The public health movement and moral statistics of the 19<sup>th</sup> century were a response to the dreadful living conditions of the working class in the era of industrialization. These conditions risked leading to social unrest and insecurity for the middle and upper class. New movements are contingent with the environment (see also Flynn (2002) for an assessment of the context of NPM).

Measurement in the public sector is often seen as an intrusion of private sector rationale in the public sector. Particularly in the 1980s, government was urged to look more like the private sector (Bouckaert

1990). The contrast of the success stories of the TQM movement with the debris of government after the economic crisis of the late 1970s presumably fuelled this opinion. Nonetheless, the history of measurement in the public sector shows a lot of public sector innovation. The budget reforms in the U.S. Ministry of Defense in the 1970s for instance were exported to the private sector (Novick 1973). Secondly, when *Adolphe Quetelet* introduced the average as a concept in the 19<sup>th</sup> century, this was a big step forward for the private sector too.

In many European countries, quantification in government is seen as an Anglo-Saxon *fremdkörper*. The current perception may be ascribed to the NPM movement that originated from the Anglo-Saxon world (see for instance Kickert 1997). Yet, in making an historical judgment, people tend to outweigh recent experiences. A landmark text of the OECD on performance management in NPM illustrates the Anglo-Saxon focus. Although Scandinavian countries too adopted performance management, it was predominantly seen as an Anglo-American management style. On the other hand, a more rule and norms-based management style was considered typical for Continental and Latin countries.

	<i>Anglo American</i>	<i>Scandinavian;</i>	<i>Continental;</i>	<i>Latin</i>
Performance Management	New Zealand, Australia, United Kingdom, Canada, United States of America, Ireland	Sweden, Denmark, Finland,	The Netherlands	
Mixed Management		Norway.	Austria, Germany and Switzerland	Belgium, France
Rule and Norm Management				Portugal Spain Greece and Italy

Table 5: Public Management Styles according to the OECD (1997)

Other 20<sup>th</sup> century movements such as the science of administration and the social indicators movement too were attributed to American inventors. In the 20<sup>th</sup> century, advancements were indeed mainly Anglo-Saxon. Nonetheless, the oversight of the movements shows that many earlier inventions were European. Lazarsfeld (1961) remarks that *'nothing is stranger than the idea often expressed by European colleagues that quantification is a U.S. export endangering their tradition. It is true that when this country (the United States of America) took over the European empirical research tradition, it did so on large scale. But the steps by which this came about are little known (p.332).'* Moreover, movements spread more easily on a global scale. Therefore, the originator of a movement becomes less of an issue in determining the nature of a movement. Rather, the transformative processes through which global movements are translated to local instances should be the prime focus for explaining the nature and the results of movements on practice (Christensen and Laegreid 2001).

A duo of interesting issues is a) why we remember a particular movement today and we do most likely not remember others, and b) why movements travel around the world, notwithstanding their contingent features. To answer these questions, we need to look at the carriers of the ideas of the movements. How are ideas conveyed through time and place? The analysis of the history of measurement in government suggests some carriers:

1. First, individuals are important. All movements have some main proponents. They symbolize the movement. Names like William Farr (Public Health Movement), Adolphe Quetelet (moral statistics), Ernst Engel (German historical school), Frederick Taylor (scientific management), Woodrow Wilson (science of administration), or the quality guru's such as Juran or Deming are emblematic for their respective movements (supra). Through these main figures, movements are identifiable in different places and time segments.

2. Individuals are important in yet another way. Personal career paths and networks make ideas flow. For instance, Jane Addams introduced the social maps of London by Charles Booth in the United States of America. The network of Settlement houses, i.e. *Toynbee Hall* in London and the *Hull House* in Chicago was a channel for the distribution of the social survey (Bulmer et al 1991). An example of a catalyst career path is the Robert McNamara's PPBS system, which was based on previous experience with Ford Corporation. Other examples are the studies of Quetelet's work by William Farr in Paris, the many American scholars that got acquainted with the German historical school while studying in Germany, or the presidency of the Taft Commission by the former director of the New York Bureau of Municipal Research, Frederic Cleveland.

3. Conferences and associations are at times a major carrier. In the 19<sup>th</sup> century, with more limited possibilities for communication, the series of high-level conferences initiated by Quetelet with participants from almost every important statistical society, were a main forum for the ideas and the methods of statisticians in particular and a broad platform for discussion in general. Associations may play an important role. The *International City/County Association* for instance has a long history in disseminating performance measurement in the local public sector (Bouckaert 1992). A more recent example is the Public Management Section (PUMA) of the OECD, which promoted NPM concepts in its member states.

4. Fourthly, written documents are important carriers for movements. (Semi-) academic texts, usually written by the main figures of a movement, are used to disseminate the ideas of the movement. These key texts are used for research, training and advocacy. While the leaders of the movement are the prophets, the key texts are the Bible. One of the key texts of the NPM movement for instance is Osborne and Gaebler's *Reinventing Government* (1992). It is well written and persuasive. Although the book is practice-oriented, it is larded with scientific argumentation. Other movements have similar key texts. The social indicator movement for instance is often traced back to Bauer's (1966) assessment of the side effects of the space program.

5. Finally, the 20<sup>th</sup> century first witnessed the exportation of ideas as a deliberate policy. The *New York Bureau of Municipal Research* intentionally exported its work to other communities through the provision of services and through contacts with agencies and officials. The PPBS system too was intentionally promoted in other countries as well as in the private sector. The same applies to NPM.

Many delegations from all over the world visited in the late 1990s the NPM champions such as New Zealand and the United Kingdom. Regularly, a country needs to subscribe to a movement in order to receive support from international organizations. Mongolia for instance adopted a New Zealand style reform that was heavily supported by the Asian Development Bank (Asian Development Bank 2004)<sup>45</sup>. Another example are the Eastern European countries that apply PPBS-like systems in their armies in a NATO context. Measurement increasingly has external, promotional purposes too. This is not surprising. Measurement has become big business. The underlying evolution is the increasing quantification of an ever-larger part of the public sector.

A common thread throughout the history of measurement is the relation between data and analysis. Some advancements mainly resulted in a supply of data. Examples are the German statistical tradition, the census taking and the social indicator movement. Other progressions were mainly on the methodology of obtaining and analyzing data. Some examples include the political arithmetic and the moral statistics. The symbiosis of data and systematic analysis within government remains a challenge. One of the problems of NPM was the mere integration of performance data, and not performance information, in the management tools. A blank is the institutionalization of methodological advancements together with new data and new management and policy tools.

*Plus ça change, plus c'est la même chose?* The historical analysis above provides evidence for some change, but usually more modest than hoped for by the initiators of change. Four trajectories of genuine change may be distinguished.

1. From ad hoc to systematic. First, there is an evolution from ad hoc measurement to more systematic measurement. Initially, peripheral actors measured government performance, mainly to influence decision makers. What and when to measure was influenced by the needs of the day. Nowadays, measurement is done on a more regular basis - often laid down in management scorecards and management information systems. ICT applications are an important factor in the recurrent supply of data (which is not necessarily information).

2. From generic to specialized. Secondly, measurement became increasingly specialized. In particular the social indicators and Evidence Based Policy movements are witnesses of an increased specialization in policy sectors (Bauer 1966; Davies, Nutley and Smith 2000). The increasing specialization of policy sectors led to a more specialized supply and demand for information within

---

<sup>45</sup> The main piece of legislation is the Public Sector Management and Finance Law. It lays down the guiding principles for fiscal management, which require that annual priorities be identified in business plans and that outputs be delivered based on agreements (or contracts). Planning is based

policy sectors. The concern of both the Social Indicators as the Evidence Based Policy movement was that this rich sectoral knowledge base is insufficiently opened up for decision makers. The gap between the increasingly specialized supply and the generalist demand of political decision-makers is an increasingly difficult issue in performance measurement.

3. From general to professionalized. A third and parallel trend is the increasing professionalization of measurement. This trend has two dimensions. On the supply side of information, professionalization implies that measurement has increasingly become a profession. There is an increasing number of measurement professionals. On the demand side of information, there is a more professional dealing with information. The way in which information is used (cf. the longlist of uses) is increasingly complicated.

4. From anecdotic to institutionalized. Increasingly, measurement got embedded in the management and policy making systems of government. For instance, in the last 50 years, considerable attempts have been made to integrate performance information in the financial information system; budgeting, accounting and audit. The increasing enactment of performance information in legislation also demonstrates the institutionalizing of performance measurement. Legislation may give a strong impulse for the demand of performance information. The GPRA is a good example of a demand shock.

---

on a 3-year, medium-term, rolling budget framework (Asian Development Bank 2004).

## 4. Literature study

### 4.1. Purpose of the literature study

Higher performance is generally a manifest objective of public sector reform. Therefore, performance measurement usually is a key facet in the content of the reforms (Pollitt and Bouckaert 1999). Performance measurement in the public sector is widespread and multi-faceted. Performance measures can be found among other things in HRM applications, in strategic plans, in budget allocation processes, in budget requests, in contracts and in budget documents (Hatry 1999). Performance measurement thus permeates many aspects of public administration practice. Likewise, performance measurement comes across many research issues in public administration theory. The versatility of performance measurement in administrative practice is reflected in the literature on the subject. The risk of this heterogeneity is the watering down of performance measurement as a distinguished -but not isolated- research topic with specific problems and solutions. We will pay particular attention to the supply and demand of information.

The literature study thus is intended to position our research vis-à-vis previous research efforts. Four questions will be addressed: what are the research questions; what is the theoretical orientation and range?; what are the methodologies and finally, what is the focus of the studies. In addition, the literature study has to substantiate the general framework. In particular by categorizing the research questions, we seek to gain insight in what supply and demand of performance information means in Public Administration research. We subsequently discuss the four research questions for the literature study. Beforehand, the methodology and the case selection is described.

### 4.2. Research method and data for the literature study

A literature study implies several choices (Bell 1993). This paragraph explicates and motivates our choices. First, the review only considers journal articles. Other sources such as books and theses are not analyzed. Although the latter two sources may give a more profound insight in answering research questions, articles provide a more up-to-date picture, which allows for the inclusion of the latest research findings in the review. In addition, it seems a reasonable assumption that the research issues in books and theses are comparable to those in articles, albeit with a different profundity. Secondly, the articles were selected along two lines: in depth and width-ways. For the in depth analysis, the contents of four journals has been considered from 1985 to 2004. The four journals are *Public Performance and Management Review*, *Public Administration Review*, *Public Money and Management and Financial Accountability and Management*. In order to surmount sub-disciplinary boundaries, the in depth analysis was complemented by a more general search using electronic

databases: *Academic Search Elite*, *Academic Search Premier* and *Business Source Premier* (through EBSCO) and the *Social Science Citation Index*. The search expression included combinations of “performance”, “measurement”, “government” and “public sector”. This search resulted in articles in several other journals (see appendix for the complete list). The third decision is the time span. Only articles published from 1985 to present are analyzed. Largely, this is an arbitrary decision. However, it is important to notice that the year 1985 goes back before the bulk of the New Public Management reforms (Light 1997). The last limiter is the publication language. Only articles that are published in English have been included. Again, this is a pragmatic choice, which is nonetheless defensible. Many European scholars publish in English journals. Moreover, the performance measurement literature is largely English. For instance, a review of the leading French journal *Politiques et Management Publique* only yielded two articles (Pallez 2000; Kopel 2001).

There are three appendices to the literature study. The first appendix is the list of articles comprised in the review (appendix 13.1.1). Secondly, the classification of the research questions according to the supply and demand framework, as well as the main conclusions of the articles are appended (appendix 13.1.2). The third appendix is the classification of the studies according to the variables we discuss in the remainder of this chapter (appendix 13.1.3).

### **4.3. What are the main research questions in performance measurement research**

#### **4.3.1. About the research question**

The first research question is about the questions that studies ask. We use the general framework to categorize the research questions. A systematic review of the studies we included in the analysis is appended (13.1). Four categories of variables that cover the majority of the research variables in performance measurement research are identified (Figure 8).

- variables describing the causes and conditions for performance measurement,
- variables describing how the performance information is used (demand side),
- variables describing how the information production process advances (supply side),
- variables describing the effects of the introduction of performance measurement on the organization, the policy sector or government as a whole?



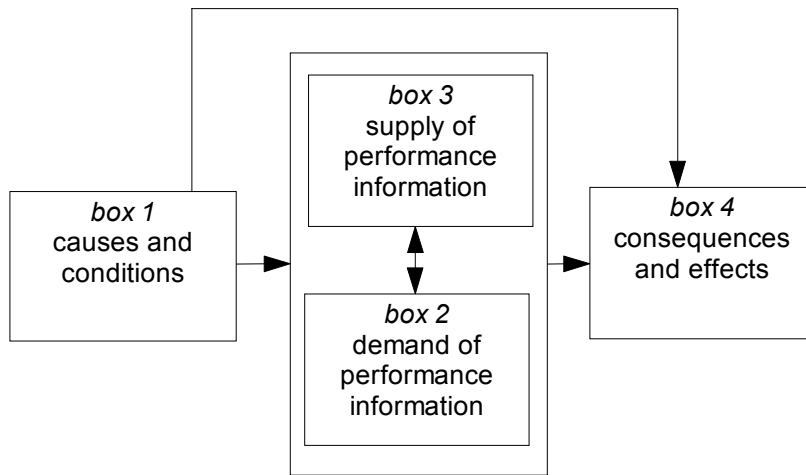


Figure 8: A categorization of variables in performance measurement research

Three sets of independent and dependent variables are distinguished. First, some studies try to establish the causes and conditions of supply and/or consumption. In addition, causes and conditions for effects of performance measurement may be addressed. Secondly, studies relate supply and consumption. At this point, the causal direction is not clear. Typically, demand determines the supply, i.e. the design of the measurement system. Which measurement system do we need in order to address a particular way of use? However, researchers frequently reverse the relationship. The research question then is how the design of a measurement system enables or distorts the use of performance information. This implies that supply becomes the independent variable and consumption the dependent variable. Finally, a third set of studies investigates on the effects of performance measurement and management in the organization and its environment.

#### 4.3.2. Findings

The research questions can be classified within three sets of independent and dependent variables. The three sets try to answer three different questions. First, the set on the causes and conditions deals with the questions why and when. Secondly, the set that relates design and use is dealing with the question how. Thirdly, the set that studies the effects of performance measurement is investigating on the question of the consequences of measuring and using performance information.

## SET 1: CAUSES AND CONDITIONS FOR PERFORMANCE MEASUREMENT

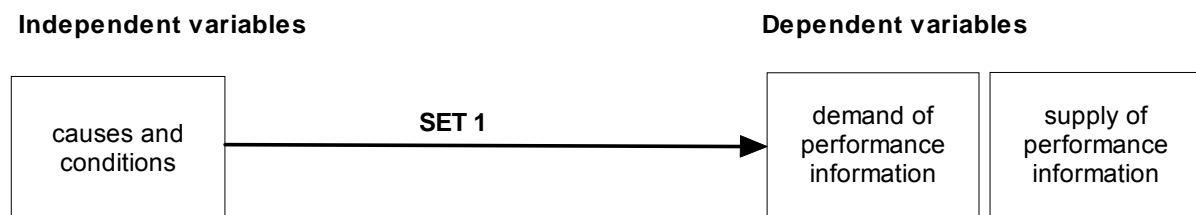


Figure 9: Set 1: causes and conditions for performance measurement

As Figure 9 shows, research questions in set 1 address the causes and conditions that explain the introduction of performance measurement. Independent variables usually relate to the organization, the organization's output, or the organization's environment.

**Organizational factors** can be found in De Lancer and Holzer (2001). They made a distinction between the adoption and the implementation of performance measurement. Rational/technocratic factors (such as having a goal orientation, committing resources, being informed about the techniques and the existence of external requirements) influence more heavily the adoption. Political/cultural factors (such as the existence of internal interest groups, and risk taking attitudes in the organization) influence more heavily the implementation stage. Behn (2002) adds a number of psychological barriers to successful adoption and implementation. He proposes a mental reorientation, which requires a new way of thinking for the many actors involved in performance measurement. Poister and Streib (1999) also stress the prevalence of organizational characteristics. They found that motivation to measure performance is predominantly internal (better decision making) and not external (legal requirements) to the organization. Finally, Wang and Berman (2000) point to the importance of central management involvement and mission orientation and to a lesser extent professional competency and resource availability as organizational prerequisites for the deployment of performance measurement.

A second explanatory variable is the **organization's output**. The central argument is that organizations with tangible outputs will more easily adopt and implement performance measurement. In these cases, the characteristics of the output are the main explanatory variable for successful measurement. Berry, Brower and Flowers (2000) find that the routine technologies facilitate a more coherent implementation of *Performance Based Accountability*. Another example is the study of Lindkvist (1996) in a large Swedish hospital. The author identifies the low measurability of services and lack of knowledge on prices and costs as two important factors explaining the failure of performance based budgeting.

A different subject relating to output as an explanatory variable is the **transferability** of private sector approaches to the public sector. The limited measurability of public sector output and the limited

insight in causal relations between output and effect are often seen as the main difference between the public and the private sector. Hedley (1998) for instance points to the difficulties of using private sector methods for performance measurement in the public sector, the most significant of which being the '*seemingly incompatibility of efficiency and effectiveness controls* (p. 257)'. Carter (1991) however compared public organizations with private monopolists (with some government control) and concluded that the public - private divide does not explain variation in the problems with performance measurement.

Thirdly, the **organization's environment** may influence performance measurement. Performance measurement takes place in organizations that operate in a societal context. A number of contexts are listed here. The spirit of the age influences the appearance of performance measurement (Bouckaert 1990; Sanderson 2001). An important positively related environmental factor is support from politicians and stakeholders (deLancer and Holzer 2001; Berman and Wang 2000; Berry Brower and Flowers 2000; Wang and Berman 2000; Broom 1995) Negative factors are resource scarcity (Reck 2001), unionization (Delancer and Holzer 2001) and the existence of a local monopoly (Linkvist 1996). No association is found between performance measurement and demographic factors such as state income and unemployment (Lee and Burns 2000) and decentralization efforts (Wang and Berman 2000). Finally, Roy and Séguin (2000) found from a case study that organizations adopt efficiency-oriented approaches because of institutional pressure. This research confirms Powell and DiMaggio's institutional isomorphism theory (1983).

SET 2: RELATING DESIGN (SUPPLY) AND USE (DEMAND)

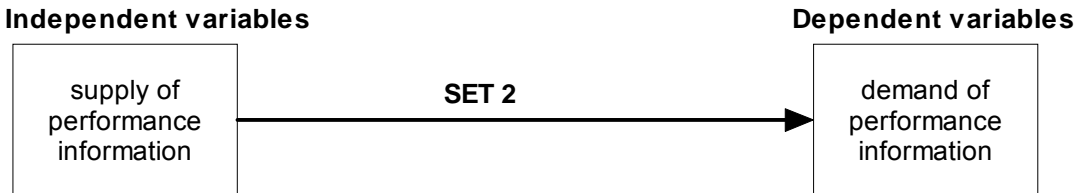


Figure 10: Set 2: supply and demand

The second set of research questions relates design parameters (the supply side) to the use of performance information. The issue is how to build performance measurement systems that supply information that is '*fit for use*'. Some issues and exemplary studies are noted below.

First, a repeated issue concerns **quality of performance measurement systems**. Bouckaert (1993a) emphasizes that performance measurement systems that serve the purposes of contemporary management should not only be technically valid, but also legitimate and functional. This requires a shift in the design of performance measurement systems. Rivenbark and Pizarella (2002) point to the importance of auditing the performance data in order to enhance its usefulness. Mol (1996) found from

a case study in the Dutch army that not so much the quality of the data caused the limited use of performance information. Mainly the selection of the indicators and the analysis of the data were problematic. In fact, the indicators chiefly referred to secondary processes in the organization. Moreover, the indicators were not related to each other. The link of performance indicators with organizational or program goals is repeatedly seen as an important quality of performance measurement systems.

Tying objectives and indicators together is often perceived as a critical success factor for successful implementation and use (Berry, Brower, and Flowers 2000; Heinrich 1999; Glaser 1991). However, one six-case study in Norway (Johnsen 1999) found that successful cases used a decoupled model, i.e. not linking performance indicators to objectives. The author proposes two explanations; i.e. ambiguity and resistance. Implementation of programs requires power, which in turn requires coalition formation and the 'overselling' of the program's potential. Performance indicators that are tightly coupled to oversold goals may be unreliable, vague and highly ambiguous. Another explanation may be that coupling indicators with goals leads to increasing environmental resistance and conflict over the goals, since the indicators make the goals more concrete.

Many studies on quality of performance measurement are based on annual reports. For instance, Boyne and Law (1991) analyzed annual reports of Welch district councils from 1981 to 1988. They concluded that the annual reports are generally of poor performance since most of the performance indicators refer to service input while broader issues concerning 'citizenship' and equity of service receive little attention. Moreover, they have not observed improvement in the annual reports over time. Other examples are the studies by Hyndman and Anderson (1995) of 57 agencies' annual reports searching for highlighted performance statements and by Rutherford (2000) on the comprehensibility, comparability and the perception of importance of indicators in 44 UK agencies' annual reports.

Next, the **management of expectations** is a noteworthy research concern. Ammons, Coe and Lombardo (2001) for instance evaluated how to make performance comparison projects more valuable for participants, based on experiences of participants in three benchmarking projects. They conclude that expectations were higher than the results of the benchmarking projects. This is not surprising. The benefits of performance measurement usually are intangible and materialize in the long term. The costs usually occur in the short term and in hard currency and (Bouckaert and Peters 2002). Therefore, expectations management is important. Benchmarking should be represented as a tool. The tool-metaphor is helpful in adjusting expectations. A tool needs to be used; just having it is not enough. A tool needs to be assembled properly and requires care and maintenance. In addition, the competing demands for cost and performance data precision, simplicity and timeliness need to be reassessed. A common denominator in the findings is the need to align expectations with what is feasible both in conceptual and practical terms. Kravchuk and Schack (1996) also stress the importance of realistic expectations when they put 'recognition and use of measures as "indicators"

only' (p. 357) as one of the main design principles in designing more formal performance measurement systems under the GPRA framework.

A third, topical issue in the design of measurement processes is the **involvement of citizens** in performance measurement. The underlying assumption is that citizen involvement ensures that *'to ensure that what is measured is what matters to citizens and that the data is not corrupted by the natural desire of officeholders to report favorable outcomes (Citizen-Driven Performance Measurement Curricular Project)*. Participation in indicator development by users may increase the perception of usefulness (deLancer 2001). Ho (2003) concluded from a survey that there is a generally positive evaluation of the concept and that practical barriers such as data availability are less of a concern. The main difficulty is a lack of communication between administrators and elected officials.

Fourthly, the **design for concrete management practices** is described. Performance appraisal is such a concrete management application. Ammons and Rodriguez (1986) studied performance appraisal for managerial performance in 170 cities. They found that 59% of the evaluations had a formal basis and 16% had a full reliance on rating scales. Moreover, 12% of the cities had no evaluation at all. Only very modest amounts of executive and staff time were devoted to the processes. England and Parle (1987) repeated the latter study for non-managerial performance appraisal. They concluded that non-managerial performance appraisal systems are more formally documented and more oriented towards skill deficiencies than managerial performance appraisal systems. Melkers and Willoughby (1998; 2001) looked for the impetus, implementation and effects of Performance Based Budgeting in states in the USA. They found that implementation was limited and that few states link between performance information and actual appropriations. PBB is as yet not effective in changing the way budget decisions are made.

Next, some studies explore the **potential of more sophisticated analysis techniques** to analyse performance information. The implicit assumption is that better analyses techniques will better meet demand and increase the use of performance information. In particular, the use of *Data Envelopment Analysis (DEA)* is studied. Nyhan and Martin (1999) for instance conclude that DEA is a useful addition to traditional ratio and regression analysis. Several inputs, outputs and outcomes can be combined in one model. Moreover, benchmarks that are more efficient can easily be identified. Worthington and Dollery (2000) applied DEA on local governments' planning and regulatory function in New South Wales. Dawson and Street (2000) evaluated the use of indices to measure unit costs in hospitals of the British National Health Service and concluded that there are considerable variations in ranks of hospitals for different indices. This has implications for the use of this data in league tables and for funding decisions.

### SET 3: EFFECTS AND CONSEQUENCES

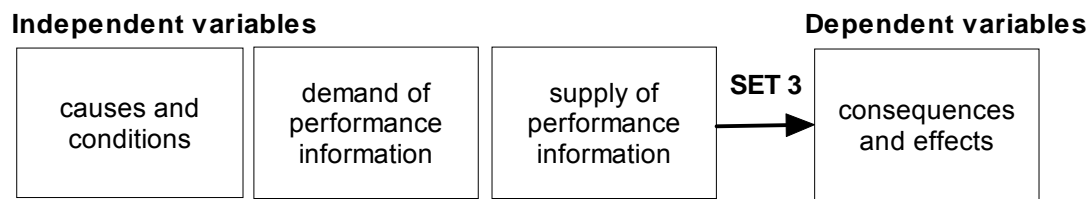


Figure 11: Set 3: consequences and effects of performance measurement

Performance measurement is not neutral. Organizations react to measurement initiatives. A third set of research questions addresses the **dysfunctional effects and consequences** for the organization of (the use of) performance information in organizations. Behn and Kant (1999) identify several pitfalls in performance contracting. The pitfalls in performance contracting, as opposed to regulatory contracting, consist of three punishment pitfalls (inhibit experimentation, encourage cost cutting rather than innovation, stifle overachievement), three revolving door pitfalls (not provide for start-up cost, inhibit symbiotic relationships, reward promises and not performance) and four complexity pitfalls (reliance on output and not outcome, distort behavior, creaming, undermine equity and fairness). Dawson and Street (2000) conclude from a case study that measurement of unit cost leads to 'creative accounting'. Heinrich (1999) evaluated the use of performance standards systems by job-training agency's program administrators. She concludes that information from the performance based contracting system is being used to make resource allocation decisions. Yet, performance measures are not strongly correlated with program goals and there is a predominance of cost-per-placement considerations, which has a negative impact on service quality. Smith (1993) infers from a study in maternity services that outcome-reporting schemes are embedded in an organization, and thus are not neutral. An over-reliance on performance indicators for control has potentially the following dysfunctional consequences; tunnel-vision, sub-optimization, myopia, convergence, gaming, ossification and misrepresentation. Other studies also describe the perverse effects or unintended consequences of measurement (Bouckaert and Balk 1991; Meyer and Gupta 1994; Grizzle 2002; van Thiel and Leeuw 2002). Interestingly, these studies almost entirely address the negative effects of performance measurement.

Finally, several scholars pointed to **compatibility problems of different uses** of performance measurement. Wiggings and Tymms (2002) studied the impact of external publication of indicators in league tables using a questionnaire. They found that English primary schools perceive their KPI systems (with league tables) as being significantly more dysfunctional than those of their Scottish counterparts are. Apparently, external publication impedes internal support. Thompson (2000) evaluated the consequences of the use of performance measurement on management practices for mid and lower level management. The research questions whether total management capacity (comprising all management levels) increases or decreases by using performance indicators. The

USA Social Security Administration (SSA) served as a case for their analysis?. The SSA has a strong use of the "big four" indicators (essentially output indicators on processing of claims). These indicators are used for pay decisions, promotions, and ranking district offices. Indeed, top management and political levels have used the measures to increase control over middle and lower level management. However, measures became a substitute for a more in depth decision-making processes. By using the measures for control purposes, the discretion needed to implement the management function at mid and lower level has eroded. The total management capacity is reduced. Moreover, because of connection of management functions, the affection of one management function results in a more than proportional reduction in total management capacity. The author concludes that there is a dual potentiality in performance measurement (i.e. better management or better control), which may not be easily combined.

Table 6 provides an overview of the clusters of research questions.

	Set 1	Set 2	Set 3
<b>Clusters of research questions.</b>	1. Organizational factors 2. Output and outcome characteristics 3. Transferability of private sector approaches 4. Organization's environment.	1. Quality of measurement systems 2. Management of expectations 3. Involvement of citizens 4. Measurement design for concrete management practices 5. Potential of more sophisticated analysis techniques	1. Dysfunctional effects and consequences 2. Compatibility of uses of performance information

Table 6: Summary table of research questions

**4.4. What is the orientation and the range of the theories used for studying performance measurement?**

**4.4.1. About the research question**

In this section, we look at the orientation and the range of the theories. The theoretical orientation of the articles refers to the theoretical paradigm, which may be positivist or interpretative. This is an important distinction in social sciences (Innes 1990). Positivist theories look for probabilistic causal laws that can be used to predict general patterns of human activity (Neuman, 1997: 63). Positivists look for universal laws. In the field of performance measurement, positivist approaches would typically put forward factors for success and failure. Interpretive theories on the other hand want to arrive at understandings and interpretations of how people create and maintain their social worlds (Neuman, 1997:68). In this tradition, reality is socially constructed. Research explores the meanings actors attach to situations. In the field of performance measurement, a typical study would theorize about how reality is constructed through the definition of indicators. Performance of a public broadcasting

company may for instance be equaled to viewing figures. Etzioni (1967) calls this phenomenon 'concept reduction'. The definition of a concept is reduced to purely to its measurable elements<sup>46</sup>.

Besides the orientation, we will also evaluate the range of the theories. This distinction is based on the insightful analysis of the role of theory in social sciences by Merton (1968). Theories can be distinguished by their intrinsic characteristics. The extreme positions of the range are the total theoretical systems on the one hand and the necessary working hypotheses that evolve in abundance during day-to-day research on the other hand.

First, middle range theories differ from daily empirical generalizations because the middle range theories consist of assumptions that make it possible to deduct testable hypotheses. An isolated proposition that describes observed similarities and connections between a set of variables could not do this. Merton considers a middle range theory as describing a phenomenon in a relatively elementary way. A good theory acquires theoretical value through its use. The usefulness of the theory may be tested by its capacity to posit new questions. These questions can be both theoretical and empirical.

Secondly, the middle range theories differ from the total theories because the former only explain one specific aspect of social behavior, social organization and change. The most prominent example of a grand theory is Parsons' functionalism. Nowadays, Luhman's social system theory is an important example of a grand theory (Brans and Rossbach 1997). Another example of a contemporary broad range theory is Giddens' structuration theory (Giddens 1986). Theories of the middle range are not necessarily deducted from an encompassing framework, but they may often fit in, in hindsight. Merton even argues that middle range theories may fit in different encompassing frameworks. Comprehensive sociological theories are sufficiently loose-knit, internally diversified and mutually overlapping that a given theory of the middle range, which has a measure of empirical confirmation, can often be subsumed under comprehensive theories, which are themselves discrepant in certain respects (Merton 1968: p. 43).

---

<sup>46</sup> Etzioni studied another example, i.e. IQ-tests. Inborn intellectual capacity is often equaled to the scores on an intelligence test.



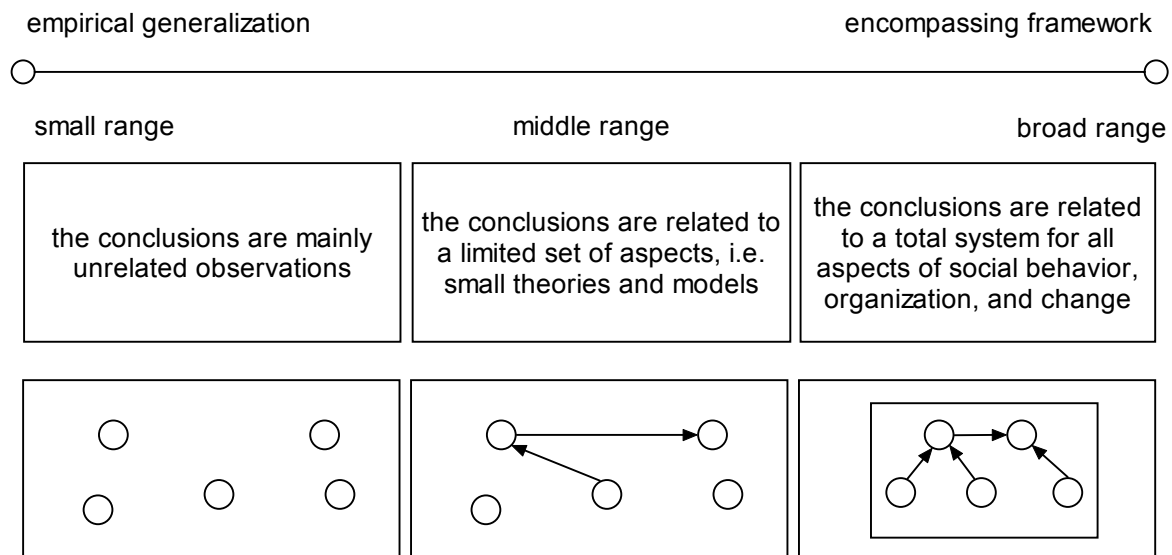


Figure 12: Small, middle and broad range theories.

Figure 12 represents the different ranges schematically and indicates the way in which we will assess the studies in this literature study. We will distinguish between the small range, the middle range and the broad range. Small range studies make observations but do not link them to each other or to explanatory contextual variables. Studies that prescribe how to do performance measurement without attributing the prescriptions to contingent factors are also considered small range. Middle range theories are positing a limited set of relations. Typically, these studies allow to put if/then/else – like statements. Additionally, middle range studies should have the potential to straightforwardly formulate new hypotheses. Thirdly, broad range theories give a complete picture for all aspects. Obviously, such theories are usually not elaborated in journal articles. We thus do not expect to find broad range frameworks in the articles. Therefore, as a proxy, we assessed whether the authors refer to these total systems of theory in the articles or not. In other words, we assess whether performance measurement explicitly is given a place within a grand theory.

A total system in which observations about every aspect of social behavior, organization, and change find their place is according to Merton not feasible yet<sup>47</sup>. Unlike the specialization in the physical

---

<sup>47</sup> According to Merton, the pursuit of a total system is based on one or more of the following misconceptions about science. First, there is the misconception that systems of thought can be developed before a great mass of basic observations has been accumulated. In the social sciences, systems often issue fully formed from the mind of one man. Progressive adaptive modification because of concerted efforts of great numbers of men is rare (p.47). A second misconception is that all cultural products existing at the same moment in history have the same degree of maturity. Compared to other sciences, social science is a relatively young science. Thirdly, social scientists sometimes misread the actual state of the physical

sciences, differentiation in social sciences yielded theoretical systems that are held to be mutually exclusive. Rather than a division in tasks, specialization in the social sciences leads to a competition between the specialisms. We will assess whether this observation is valid for performance measurement studies too. Figure 13 summarizes the characteristics of Middle Range Theories.

1. Middle range theories consist of limited sets of assumptions from which specific hypotheses are logically derived and confirmed by empirical investigation.
2. Middle range theories do not remain separate, but are consolidated into wider networks of theory. There is a hierarchy of theories. Theories that are higher up in the hierarchy explain more phenomena than lower theories.
3. Middle range theories are sufficiently abstract to deal with different spheres of social behavior and social structure, so that they transcend sheer description or empirical generalization.
4. Middle range theories cut across the distinction between micro-sociological problems and macro-sociological problems - i.e. the experimental studies of small groups and the comparative analysis of specified aspects of social structure. Macro subjects do not necessarily need macro theories and micro subjects do not inevitably require micro theories.
5. Middle range theories have total systems of theories as general theoretical orientations, rather than rigorous and tight-knit systems. In this study, the general theoretical orientations are neo-institutionalism as a theory of organization and knowledge utilization as a theory of information.
6. Middle range theories, as a result, are often consonant with a variety of total systems.
7. Middle range theories are in line with classical theoretical formulations such as Durkheim and Weber. Merton also refers to Francis Bacon (middle axioms in science)<sup>48</sup>, John Stuart Mill, Karl Mannheim ("principae media") who formulated a hierarchy in theoretical reach of theories, and advocated the middle range.
8. Middle range theories involve the specification of ignorance. It expressly recognizes what must still be understood?.

Figure 13: Characteristics of Middle Range Theories (adapted from Merton 1968).

---

sciences. Even in the physical sciences, a univocal total system is not established. Fourthly, Merton observes utilitarian pressures for total systems. In modern western societies, large demands are made of social sciences - by policy makers, by reformers and reactionaries, by businessman and government man. The promise of social science should be put into proportion. *"The urgency or immensity of a practical social problem does not ensure its immediate solution. At any given moment, scientists are close to the solutions of some problems and remote from others. It must be remembered that necessity is only the mother of invention; socially accumulated knowledge is its father. (p.50)"*

<sup>48</sup> Bacon in turn refers to Plato who stated, "that particulars are infinite, and the higher generalities give no sufficient direction" (Merton 1968; p.56).

#### 4.4.2. Findings

The vast majority of the studies has a positivist theoretical orientation. Only one out of ten has interpretative aspects. A recent example is Mausolf's study on learning from feedback of performance measurement information (2004). He finds that every phase in the learning process yields new information, and thus new interpretations. Information thus is permanently redefined, rather than imputed in the beginning of the process. Berry, Brower and Flowers (2000) suggest that normative and cognitive institutionalism may explain performance budgeting better than agency theory. DeLancer and Holzer (2001) point to the importance of risk taking attitudes in the implementation of performance measurement. The vast majority of the studies however, is univocally positivist. They search for the rules that govern performance measurement.

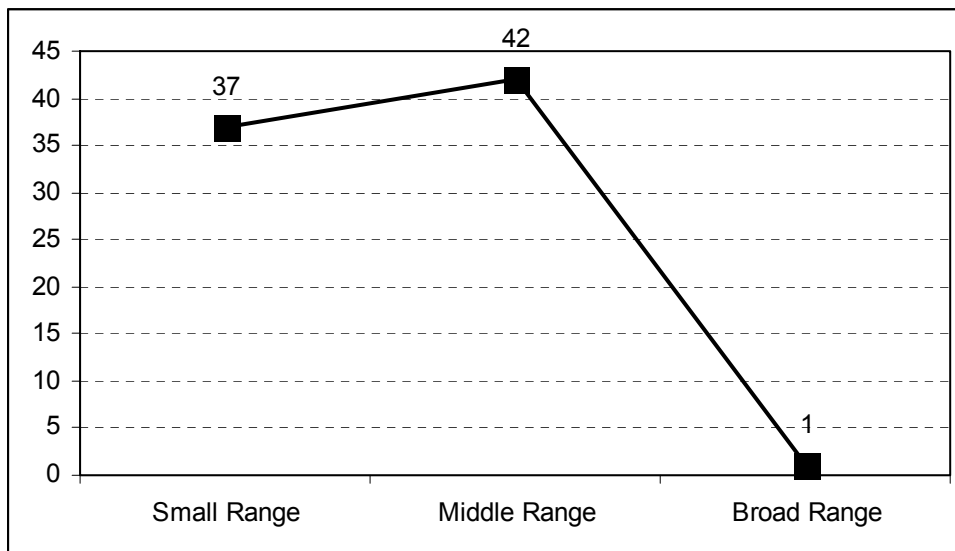


Figure 14: Range of the performance measurement studies

The range of the set of studies is represented in Figure 14. Studies focus almost exclusively on the small and middle range. They observe and posit relations between limited sets of variables. One study refers to a broad ranging theory, i.e. structuralism. Modell (2004) describes how performance measurement uses may be seen as myths. The '80s performance measurement myth involved financial control. The '90s myth holds notions of goal directed and multidimensional, and shared models. The performance myth pivoting around the supremacy of goal-directed, multidimensional measurement models gradually replaced the myth that public service provision may be improved by heavy reliance on financial control. Most of the studies however focus on the middle range and the small range. For exemplary purposes, we provide a typical example of both. A small range focus is found in Nyhan and Martin (1999) who assess that performance measurement will become increasingly important due to GPRA, National Performance Review, community benchmarking and Government Accounting Standards Board's Service Efforts and Accomplishments (SEA) project. They

assert that improved analytical tools such as Data Envelopment Analysis in addition to ratio analysis and regression analysis may be useful. Although this is undoubtedly a relevant assertion, it has limited theoretical implications. Brown and Pyers (1988) provide a middle range example. They note that the inclusion of performance measures in the financial reporting is said to enhance the value of these reports for the citizenry. Yet, they found that behavioral issues will be the barriers, not technical ones. The latter statement links variables, and allows the positing of new hypotheses.

## 4.5. What are the methods used for studying performance measurement?

### 4.5.1. About the research question

Different methods may be used for studying performance measurement. A first feature of a methodology is its obtrusiveness. Unobtrusive methods do not involve direct elicitation of information (Webb et al. 1966). Non-participant observation and documentary sources are examples of unobtrusive methods. Rather than invoking information through interviews or surveys, they study the physical traces of behavior (Lee 2000). After the Hawthorne effects, it became increasingly clear that social science measurement might be artefactual (Mayo 1933). The methodological choice between asking about performance measurement and studying traces of performance measurement has significant consequences. Secondly, methodologies can observe a phenomenon in a large or a small number of observations. Large N methods are usually quantitative and rely on statistical analysis techniques. Small N methods are usually qualitative and require more interpretation on behalf of the researcher.

We distinguish between four methods based on the difference between large N and small N methodologies and the difference between obtrusive and unobtrusive methods. Table 7 represents the classification. Firstly, there are case studies based on interviews that are obtrusive and small in number. Secondly, case studies based on documents are small in number, but non-obtrusive. Thirdly, surveys are obtrusive and large-number. Fourthly, document analysis usually also has many observations, but is non-obtrusive. Additionally, we include literature studies as a fifth methodology. It is a non-empirical way of assessing performance measurement.

	<i>Small N</i>	<i>Large N</i>
Obtrusive	usually case study with interview and questionnaires	usually surveys
Unobtrusive	usually historical research	usually document analysis

Table 7: A classification of research methods

### 4.5.2. Findings

Figure 15 gives the frequencies for the methodologies in the performance measurement articles. Most frequently, they are case studies. A case study is an application of different research strategies

(document analysis, interview, historical analysis, ...) on a single or a limited set of observations. Six studies mention single case studies. The vast majority however are multiple case studies. Mostly, the observations are organizations. Yet, other types of cases are studied. Dawson and Street (2000) compared indices of performance indicators in different sectors. The units of observation of Rivenbark and Pizzarella (2002) were local audits. Sanderson's case (2001) was a reform package, i.e. the Best Value Framework in the UK.

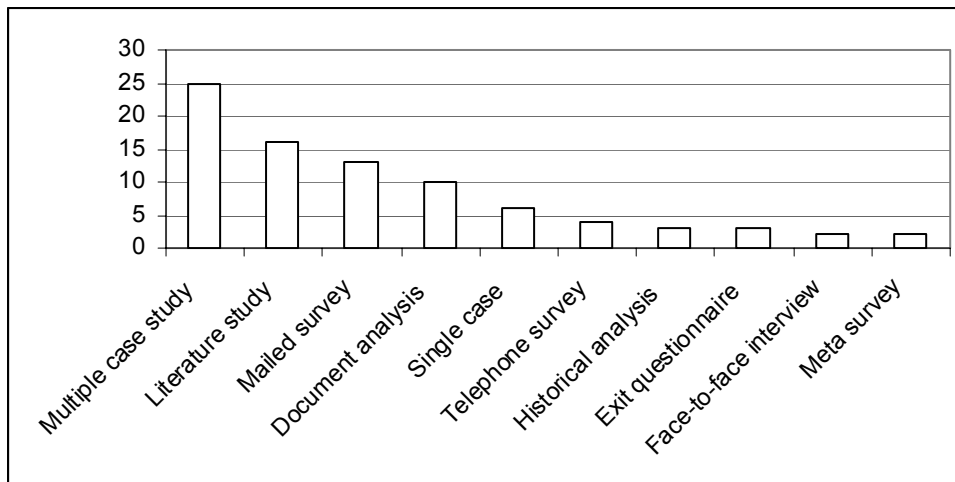


Figure 15: Methodologies of performance measurement articles

The second-most popular methodology are literature studies. Note that we define this category broadly. These are literature-based studies, not studies of the literature. The latter difference is most obvious in the conclusions of the studies. The conclusions deal with the theory and practice of performance measurement - not about the literature. The argumentation and inferences are supported by reference to the literature and not by new empirical material. Thus, this category includes mainly explorative articles. In fact, a literature study *pur sang* (a study about performance measurement literature) is not found in the set of articles.

Surveys are usually mailed surveys combined with a telephone follow-up. Most of the surveys are of local governments in the USA<sup>49</sup>. Three mailed surveys are of European origin. Boyne and Gould-Williams (2003) assessed whether planning (including target setting) affects (perceived) performance. England and Parle (1987) studied performance appraisal practices for non-managerial performance and how it differs from managerial performance appraisal. More recently, Ter Bogt (2004) published a study on the use of performance information by politicians. Survey methodology is still used in two

<sup>49</sup> Some exemplary studies are Ammons and Rodriguez 1986; O'Toole and Stipak 1988; Melkers and Willoughby 1998; Poister and Streib 1999; Berman 2000;

other ways. First, some authors perform a meta survey – surveying surveys. Lee and Burns (2000) for instance looked at a series of surveys of budget officers in 1990, 1995, 1970 and 1975. A particular approach to surveying are exit questionnaires of participants in training or Masters programs (e.g. Ammons, Coe and Lombardo 2001; Lawton, McKeivitt and Millar 2000). The infrequent face-to-face interviews generally complement mailed surveys.

Finally, some studies dig into the history of measurement. In the set of articles, three studies of this kind are included. Bouckaert (1990) looked at the history of the productivity movement in the USA. He described how the spirit of the age does influence productivity concerns, including the main methodological shifts. The time span was roughly the 20<sup>th</sup> century. Williams (2004) studied the history of measurement in the USA until 1930. He assessed that the main pivot for the development of measurement is the New York Bureau for Municipal Research. In another article, the activities of the NYBMR are studied in more depth (2003). There are no European oversight studies of the history of measurement in the public sector.

Some studies employ different methodologies. Four out of 81 studies reported to have used two different methodologies (Wang 2002; Melkers and Willoughby 2001; McKeivitt and Lawton 1996 and Lawton, McKeivitt and Millar 2000). One study used three methodologies (Berry, Brower, and Flowers 2000).

Figure 16 regroups the methodologies following the obtrusive character and the number of observations. The most popular methodologies are small N obtrusive - generally case studies. All but one study with interpretative elements (de Lancer Julnes and Holzer 2001) use a small N obtrusive methodology, or a literature study. Secondly, twenty-two studies have a large N, and an obtrusive character. These are mainly the mailed surveys. The third methodology is small N unobtrusive. These are the historical analyses. Notwithstanding the subject matter, which is in essence about quantitative measurement, many researchers opt for qualitative research. The small N methodologies and the literature studies total 63 % of the methodologies. Finally, 10 studies use unobtrusive large N. These are document analyses or analyses of statistical datasets.

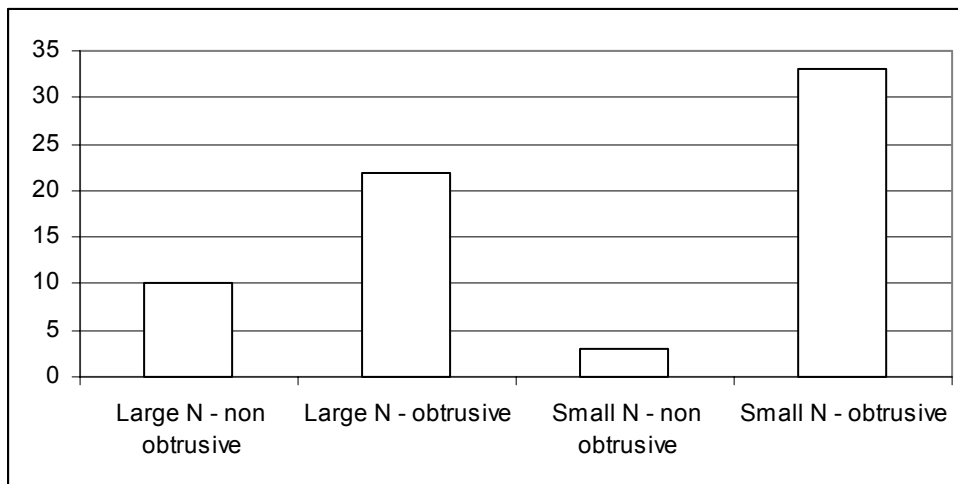


Figure 16: Obtrusiveness and No of observations of performance measurement studies

#### 4.6. Is the focus on the citizen, politics, or the administration and does it change over time?

##### 4.6.1. About the research question

Finally, we will assess whether the focus of the articles is on citizens, politicians or administrators. The politics administration dichotomy is a classic in public administration. In essence, when Woodrow Wilson created the dichotomy in the beginning of the century, it was a reaction against corruption and inertia due to political interference. Politics got back half a century later. The bureaucrats were said to run the nation(s) through technocratic devices (Fischer 1990). The bureaucrat with his/her street level discretion held the true power (Lipsky 1980). Thirdly, citizen participation literature acknowledges citizens as co-producers of public goods. Citizen participation can be more or less far-reaching (Arnstein 1969). In recent years however, citizen involvement has been promoted more. Fisher (1993) for instance asserts that the public has crucial information that can increase the quality of the process<sup>50</sup>. In performance measurement too, there is an increasing interest in citizen involvement<sup>51</sup>. . Citizen involvement is presented as something new. We will study this assertion by assessing whether the focus changes over time. We will use four time bands; 1985-1989 (the period before the NPM), 1990-1994 (early NPM), 1995-1999 (late NPM), 2000-2004 (post NPM). Before we describe the results of the literature review, a short note on the research method and the data of this literature study is provided.

<sup>50</sup> Citizen participation is not without problems. One of the most dragging issues both in practice and theory is the tension between the ideal of democracy and the reality of low participation in these initiatives (Innes and Booher 2004).

<sup>51</sup> Kloby and Kim (2004) list 30 recent publications on performance measurement and citizen participation.

## 4.6.2. Findings

Table 8 represents the focus of the performance measurement studies. The rows represent the focus. Interestingly, there is a build-up in the focus. All the studies include attention for the administrative system. These are for instance ministries, agencies, and city managers. Forty-eight studies (58%) stop here. Twenty studies (24%) additionally include attention for the political system. In addition to the political focus, fourteen studies (17%) focus on the citizen and his role in performance measurement.

The columns include four periods. The first period ('85-'89) is the pre-NPM era. There are relatively few studies. The studies focus mainly on the administration. Three studies however focus on the citizen and the politician (Mayston 1995, Ammons and Rodriguez 1996; Brown and Pyers 1988). The number of publications did not increase significantly in the second period ('90-'94). This was the advent of the NPM. In this incubation period, NPM was mainly advocated in New Zealand and the United Kingdom. The focus of performance measurement researchers was predominantly on the administration. The third period ('95-'99) was the globalization of the NPM-rhetoric. The number of articles on performance measurement increased. The boom in publications came about in the fourth period ('00-'04) when the evaluations of the NPM were brought out. Moreover, the proportion of the articles that focused on politics, and politics and citizens increased significantly. This changing focus probably reflects one the main criticisms of the NPM; the disregard of political rationality and the societal function of the government.

Focus of performance measurement studies	5 year time period				Total
	'85- '89	'90- '94	'95-'99	'00-'04	
Administration	6	8	10	24	48
Administration and politics	0	1	4	15	20
Administration and politics and citizens	3	1	1	8	13
Total	9	10	15	47	81

Table 8: Focus of performance measurement studies

## 4.7. Discussion and conclusions

The review of the studies on performance measurement demonstrates the many aspects of research on performance measurement. A few discussion, relevant for further research, points may be put forward.

First, studies prescribing optimal performance measurement systems (i.e. the supply side) usually do not relate the prescriptions to a specific demand/purpose. Nonetheless, this may be important (Behn 2003). Performance information for internal learning purposes has fundamentally different characteristics than performance information for performance contracts. This will reflect on the measurement system. Thus, the research question "how to design measurement systems that are 'fit



for use'?" should be supplemented with the question "fit for which use?" Likewise, different effects and (the lack of) results of performance measurement probably find their origin in the use and the design of the measurement system. We will further study the link between design of a measurement system and the use of performance information in chapter 8.

Secondly, the (in)compatibility of different uses is an important issue - both for theoretical and practical purposes. Is it feasible to implement a performance measurement system that intends to increase both control and organizational learning? In Halachmi's (2002) words: "*This raises the question whether the same schemes that measure performance as a means of improving accountability are as suitable for facilitating better performance or greater value for money? (p.230)*" The author draws a parallel with Mintzberg's (1994) analysis of strategic thinking and planning. Mintzberg (p.108) stated that strategic thinking involves intuition and creativity, using tacit knowledge and 'soft' data. Planning, on the other hand, requires rational and systematic analysis using 'hard' data. Ceteris paribus, performance measurement for performance improvement involves the analysis and synthesis of data that leaves considerable room for interpretation. Performance measurement for accountability on the other hand requires incontestable data that is consistent and comparable over time and organizations. Thus, the diversification of use in the study of performance measurement is important. Diversification in utilization may also have an impact on the effects of performance measurement as well as on the design of the measurement system. The systematic breakdown of utilization is a crucial point of departure in chapters 8 and 9.

Thirdly, studies of the effects of performance measurement provide appealing concepts, often termed measurement pathologies and perverse effects. They accurately point to the fact that performance measurement is not a neutral device and may have negative consequences. Yet, these studies are confronted with two deficiencies. First, they are not empirically tested. They are usually based on scattered experiences with public administration practice, including press coverage. As a result, these studies are mainly explorative. We will assess the occurrence of the effects in the Ministry of the Flemish Community in chapter 9. Secondly, the studies have the character of laundry lists. They do not search for underlying classifications and mechanisms. Chapter 9 puts forward goal displacement as the underlying mechanism of many dysfunctional effects.

Fourthly, the studies of the effects of performance measurement may be strengthened when performance measurement is seen as a multi-dimensional concept. It should be clear whether the perverse effects are caused by the performance measurement itself, or a specific use of the performance information. Moreover, when it is clear which use of performance measurement is the independent variable, the conclusions may be easily extrapolated to performance measurement as a whole. Although performance measurement is multi dimensional, it is evaluated based on one aspect, instead of the whole of the system.

Fifthly, although the whole array of methodologies are used, small N methodologies and literature studies are predominant. This may be because of the research infrastructure at universities. Large N research is usually more expensive. However, the insight that performance measurement in government is a complex multi-faceted operation may also explain the focus on more qualitative research approaches. Probably, the main step forward for methodologies lies in the combination of qualitative and quantitative approaches. Large N research such as surveys may scratch the surface and point to aspects that require further investigation. We will use different methods. The complementarities of the methods is described in chapter 5.

Sixthly, the interpretative theoretical approach is not utilized very often in performance measurement research. We do not make a fundamental choice between interpretative and positivist approach. We follow a pragmatic middle range approach. Yet, we believe that performance measurement research may benefit from the infusion of more interpretative elements. A source of inspiration may be theories on knowledge utilization. They have a longer track record in combining positivist and interpretative elements (Weiss 1977; Rich 1997; Rich and Oh 2000). It seems that traditional performance measurement research may benefit from theories on knowledge utilization. It may strengthen the theoretical foundations of performance measurement. It may also give performance information its place amid other sources of information. Decision makers usually have several other information sources besides performance information. This information width, and the relation of performance information to other sources of information may be a fruitful addition for performance measurement research. Chapter 6 applies the model of Beyer and Trice (1982) on performance measurement.

Seventhly, the theoretical course of performance measurement research is on the middle and small range. We do not attempt to alter this direction in this study and will use theories of the middle range. The middle range may be the right theoretical position. Herewith, we follow the approach for Comparative Public Administration that is described by Brans (2003), but which seems appropriate for performance measurement too. It is stated, *“there is no grand theory to answer all (...) central questions of both public administration and Public Administration.<sup>52</sup> (...) The questions are better approached from general frameworks in which the core dependent variables are agreed upon, the collection of the required original and secondary evidence is feasible and the theoretical exploration of interconnectedness between variables is meaningful. (Brans 2003: p.435)”* Jreisat (2005) also argues that *“future research has to utilize middle-range concepts to enhance the specificity and relevance of findings. (...) Middle-range models are efficient tools for applying evidence to a few administrative aspects at a time, for linking concepts to each other, and for providing balance to the abstract and the concrete in the formulation of hypotheses. (p238)”* He also points to the fact that middle range theories do not necessarily have to take sides in paradigmatic debates. *“Whether research is slanted toward*

---

<sup>52</sup> Public Administration in upper case refers to the academic discipline, while public administration in lower case refers to the practice.

*administrative structures and patterned behaviors or directed toward institutional functions and priorities, middle-range research can ensure the relevance of comparative analysis of national administrative systems (p. 238)."*

Middle range theories are particularly fit for Public Administration for several reasons.

First, Public Administration is a multidisciplinary science. Adjoining perspectives are amongst others social psychology, economics and law. The diversity is both its greatest strength and most serious limitation (Denhardt 1990). On the one hand, there is an unusual richness because of the variety of perspectives. On the other hand, the field often lacks a sense of identity. Many even question whether it is possible to speak of building a coherent and integrated Public Administration theory (Dubnick 1999). Middle range theories allow for a continuation of being able to borrow from a range of perspectives. The sense of identity will be obtained through unifying the subject of study (the phenomena and research questions), rather than through a unified theory. Public Administration theory in its current state is middle range theory.

Second, Public Administration is a relatively young science. Public administration practices can be found in all times (Hood 2000). The founding fathers of the modern discipline however, are traced back to the beginning of the 20<sup>th</sup> century. Woodrow Wilson, Frank J. Goodnow and Max Weber (Shafritz, Hydes and Parks 2004). As Merton argues for sociology, it may be too early for Public Administration too to have attained encompassing theories.

Thirdly, public administration is practice-oriented research (Denhardt 2000). The engagement of both practitioners and academics in societies such as the American Society for Public Administration exemplifies this point. The distance between the researcher and the research object is relatively small for a number of reasons. First, unlike for instance historical research, Public Administration mainly is concerned with public administrations today. Secondly, unlike physical sciences, Public Administration studies social systems and thus has to consider the human factor. Research methods are generally more obtrusive than in other disciplines outside the social sciences. Thirdly, practitioners not only are involved as respondents for research; often they are commissioners of research. Many research projects are action oriented. They are expected to yield results that can be applied in practice. Middle range theories may for now be better apt for addressing concrete issues of day-to-day public administration.

Finally, the changing focus -from administration to politics and citizens- points to the contingent nature of performance measurement and performance measurement research. After the somewhat narrow concentration on public management in the nineties, public sector reform shifted attention towards the

relation between government and the society (Pollitt and Bouckaert 2004). Western governments were (and are) plagued by low trust of their citizens in government. Performance measurement is no longer seen as merely an administrative tool. It needs to be a political and a societal tool too. The changing focus of performance measurement research reflects this.

## PART TWO: EMPIRICAL RESEARCH

Part two holds the empirical research findings of this study. First, we ask why organizations are measuring. Secondly, we study political demand and administrative supply through an analysis of parliamentary questions. Thirdly, we address the issue of fitting design characteristics of the performance measurement system to uses of performance information. Finally, the effects of performance information on organizations are studied. Each of the chapters is preceded by a summary outline of the chapter. First, we describe the methodology of the study.



## 5. Methodology

This study uses multiple methodologies. The literature study showed that a multi-methodological approach is the exception rather than the rule. Yet, it has several advantages. The main benefit is that it allows for a good fit between the research questions and the methodology. This is in line with Yin's (1994) assertion that the purpose of the research method depends on the concrete research context, instead of being ordered hierarchically (i.e. cases in the exploratory phase, surveys in the descriptive phase, and experiments in the explanatory phase) (p.3). In addition, there is the possibility of triangulation between methodologies. The research issues in chapter 7 hold a more detailed methodological description. Here, we describe the four methodologies we used for answering the research questions. We will describe the interviews in chronological order of application. First, we conducted unstructured interviews. Next, we performed a document analysis. Then we sent a mail-survey. Finally, we conducted interviews again. At this point, the interviews were semi-structured. For every method, we provide a definition, the purpose and the case selection.

### 5.1. Unstructured interviews

#### 5.1.1. Definition

Unstructured interviews only establish the topic of the interview. Unstructured interviews tend not to use prepared questionnaires or interview schedules; rather they will have a number of themes or issues that they aim to explore. The wording of the question is at the discretion of the interviewer who has to construct the questions to suit the particular respondent and the flow of conversation in the interview. Another term for unstructured interviews is the informal conversational interview (Weiss 1998). Unstructured interviews rest on an interactionist interview relationship (Silverman 1993). The interviewer creates the interview context and interviewee complies or resists the definition of the situation. Denzin (1970) calls an unstructured interview an observational encounter (p133).

Yin (1994) makes a distinction between interviews with an open-ended nature and more focused interviews. Interviewers that apply an open-ended method may look for the respondent's view on the facts as well as the respondent's opinions. Yet, in some cases, the respondent is encouraged to give his or her own insights. The researcher uses these insights for further inquiry. The respondent's role may be considered one of an *'informant'* rather than a respondent. Focused interview on the other hand only take up a short period of time – an hour for example. The interview remains conversational in nature, but the point of departure is a priori established. Our approach combines aspects of both. On the one hand, we limit the interview in time (approx. 1.5 to 2 hours). Yet, the respondents are to a certain extent informants since the main purpose was to define the research subject.

### 5.1.2. Purpose

Researchers use unstructured interviews because they want to *understand* rather than to *explain* (Fontana and Frey 1994). They are particularly useful when the boundaries between the phenomenon and the context are not evident. This fully applies to the study of performance measurement. Performance measurement is entangled in administrative processes. Often people confuse performance measurement and performance management (Hatry 2002). In particular, the confusion between performance measurement and performance related pay is tenacious. Even if they make a distinction, performance measurement comes in many guises. Some equal performance measurement with management scorecards. Others relate it to policy indicators. Many people attach different meanings to performance measurement. Bouckaert and Balk (1991) even consider it a measurement pathology. They call it the pollution disease. This context is a case in point for undertaking unstructured interviews. We first have to understand the phenomenon -the design parameters and the uses of performance information- before we can explain.

The characteristics of an unstructured interview support this purpose. Nichols (1995) identifies several advantages of unstructured interviews. First, the interviewer can build rapport with respondent and the respondents can answer in their own words. This is amongst others important for designing survey questions. Secondly, the nature of the response is not limited which results in a richness of data. Thirdly, it allows for questions that are more complex. For example, an unstructured interview can confidently address questions about the historical development of the performance measurement system in the organization. Finally, meanings and attitudes will more easily emerge from an interactive and unstructured interview setting.

Nichols also lists some limitations of unstructured interviews (1995). First, there is a lack of standardization, which may inhibit comparability of data from different respondents. Secondly, the answers are difficult to analyze. Therefore, there is a considerable potential for interviewer bias. These objections appear to be less of a problem in this stage of the research. Our purpose was to develop a conceptual framework. We aimed to identify a range of design parameters and ways of utilization of performance information. We thus confronted literature with the empery from the interviews. Analysis, in the sense of searching for explanations and comparison, was less of a concern. The third limitation is that it depends on the ability of respondents to express themselves. We assume that this is less of an issue in a professional context whit highly-skilled respondents.

In this study, open-ended interviews have been the first step. The purpose of the first round of data gathering was twofold. The first objective was to substantiate the design parameters of performance measurement on the one hand and the uses of performance measurement on the other hand. The information was written down in a research report that was the basis for the general framework (Van Dooren, Sterck and Bouckaert 2003). Moreover, the information also served to study research



question 7.3; what are the design parameters for different uses of performance information. Secondly, we gained insight not only in the nature, but also in the performance terminology within the Ministry of the Flemish Community. This was a necessary condition for developing the survey.

### 5.1.3. Case selection and instruments in this study

Field researchers may employ four types of sampling (used as a synonym for case selection) (Babbie 1975). Quota samples will include observations in all the relevant categories of the universe, e.g. leader and non-leader, autonomous or not, radical and moderate, believers and non-believers, etcetera. Snowball samples include the observations where previous observations point at. Deviant samples include the cases that deviate from the regular patterns of attitudes and behaviors. Babbie notes that the aforementioned sampling techniques are less common. Mostly, sampling in field research is purposive (1975 p. 202). The researcher selects the sample of observations that he or she believes will 'yield the most comprehensive understanding of the subject of study, based on the intuitive "feel" for the subject that comes from extended observation and reflection (p202)'.

Our sample was purposive, but not based on our intuitive feel. We wanted to have 'measurement rich' cases. We assumed that we would learn the most for developing a framework for performance measurement from those cases that have experiences with measurement. Our cases were embedded in the Ministry of the Flemish Community. We used the formal hierarchy for the selection of the cases (Figure 17). We addressed the upper level of the Ministry of the Flemish Community, i.e. the 7 departments. We asked the Secretaries General of the department and the Directors General of the Administration to identify the measurement practices. We interviewed 27 civil servants in 17 sections of the Ministry of the Flemish Community. The sections were both in line and support functions.

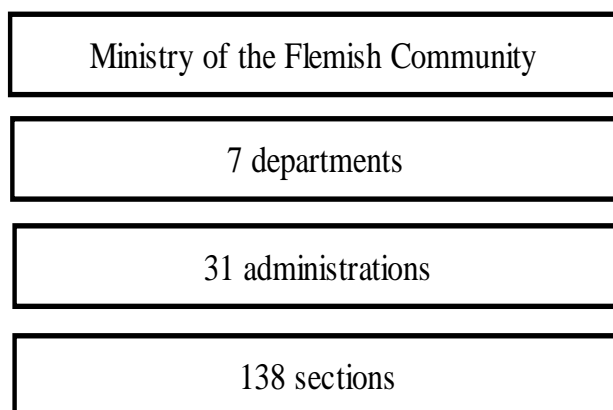


Figure 17: Organizational chart of the Ministry of the Flemish Community

Unstructured interviews require a limited amount of previous instrumentation. We only established the theme of the interview. We invited the respondents to discuss on issues about the history of

measurement in the organization, the implementation, the indicators, the analysis of the indicators, and the use of the information.

## 5.2. Content analysis

### 5.2.1. Definition

Holsti (1969) gives a wide-ranging definition of content analysis as "any technique for making inferences by objectively and systematically identifying specified characteristics of messages" (p. 14). It thus involves establishing categories and then counting the number of instances when those categories are used in a particular item of text (Silverman 1993; Krippendorf 1980; 2004). Content analysis (sometimes called textual analysis) is mainly developed in the field of mass communication (Silverman 1993; Berelman 1952). Yet, content analysis has spread out in different fields. Political scientists also applied the technique. For instance, the *Manifesto Research Group* assesses policy positions of political parties based on content analysis of party manifestos (Budge et al. 1987).

### 5.2.2. Purpose

Content analysis may be used for the study of virtually any form of communication (Babbie 1975). There are several advantages of content analysis. Probably the greatest advantage is its economy in terms of money and time. Only access to the material to code is required. We used content analysis to answer the question on political demand and administrative supply. This is the only research question with an empirical foundation that necessarily needs to be larger than the Ministry of the Flemish Community. It has to include the political arena as well as autonomous agencies. As we will argue, we found the proceedings of parliamentary questions a good proxy for the concepts we studied. Besides this pragmatic argument, there are other benefits of content analysis. First, it allows for both quantitative and qualitative analysis. Our study is mainly quantitative. Yet, we included a qualitative assessment of the reasons that ministers provide for not being able to provide the information the MP asked for. Secondly, content analysis is replicable in time and place for as long as comparable documents are available. Thirdly, it is an unobtrusive method (Hood and Dunsire 1981; Lee 2000). Social desirability does not influence the results.

Obviously, content analysis has some disadvantages too (Babbie 1975; Manning and Cullum-Swan 1994). First, content analysis looks at communications that are recorded. In many cases, recorded communications do only modestly concur with oral communication. We will look for performance information in parliamentary questions. Because of its quantitative nature, this bit of information is the least likely to get lost due to the recording. Secondly, content analysis strongly reduces the concept under study to its appearance in text. Validity thus may be an issue that needs special attention. We

will argue in chapter 7 that quantitative information in parliamentary questions is a good proxy for the concepts administrative supply and political demand.

### 5.2.3. Case selection and instruments in this study

Hood and Dunsire (1981) point to the relevance of parliamentary questions for studying political salience. We used parliamentary questions as proxy of administrative supply and political demand. The case selection was stratified. We selected 30 parliamentary questions at random for 9 policy sectors (totaling 270 questions) from different bulletins of parliamentary questions. More details on the case selection and instrumentation is provided in chapter 7.

## 5.3. Survey (questionnaire)

### 5.3.1. Definition

The survey methodology probably requires less definition. In general, it is a system for collecting information from or about people to describe, compare, or explain their knowledge, attitudes, and behavior (Fink 2003). Typically, there are two kinds of surveys (Babbie 1975). When a researcher administers the question, the survey is termed a structured interview. When the respondent administers the questions, the survey is called a questionnaire. We applied the latter.

### 5.3.2. Purpose

Yin (1994) refers to a basic categorization of research questions: “who”, “what”, “where”, “how” and “why”. The “what” question may be a qualitative description or a quantitative assessment (“how much”, “how many”). Surveys potentially can play a role for all research questions. Yet, they are usually better suited to answer the questions “what”, “how much” and “how many”. As a result, they allow for a generalization of findings to a population. Case study research has more trouble in checking representativeness (Miles and Huberman, 1994).

The first advantage of mailed questionnaires is its cost. Compared to face-to-face interview, it is considerably less expensive. Secondly, unlike telephone surveys, it can include more complex questions designs (Kumar 1996). For instance, scales that are more sophisticated are applicable. In addition, questionnaires may be longer than telephone surveys. The multidimensionality of our subject requires a relatively long set of questions. Thirdly, mail surveys allow the respondent to answer when it is suitable, rather than when the researcher gets in touch for a telephone or a face-to-face interview. In particular, for our population, section managers with usually a busy agenda, this is an important aspect. We sent out a paper version in addition to the web version of the questionnaire in order to

accommodate the respondent's time schedule. Some respondents may have wanted to fill out the questionnaire at spare moments during the day. This is impossible with a web survey since there needs to be an internet connection and the survey needs to be filled out in one go. A related advantage in comparison with interviews is that the respondent can think about the responses instead of answering off the top of the head (Weiss 1998: p. 156).

Methodologists also see disadvantages of mailed surveys (Weiss 1998; Babbie 1975). First, in some populations, response rates to mail surveys are often too small to be useful. The typical instance is a population of lower socio-economic strata. Our population clearly is not in this case. Still response rate will remain an issue. Section managers regularly receive questionnaires. The almost perfect penetration of internet and e-mail in that population – with virtually no mailing costs - has reinforced this number. The response burden for section managers is a topic of growing concern (Administratie Planning en Statistiek 2003). Our study obtained an adequate response rate of 48.3%. Secondly, mailed questionnaires usually do not reveal who in reality completed the survey. We address this issue by including a question asking for the grade of the person who fills out the survey. Thirdly, mailed surveys are mainly limited to closed questions. Superficiality is inherent to survey research in general (Babbie 1975; p.277). It is difficult to make questions entirely unambiguous. Good pre-testing of the questions should to a large extent remedy this difficulty.

### 5.3.3. Case selection and instruments in this study

The population for the survey is the 137 sections of the Ministry of the Flemish Community (see Figure 17). We did not need sampling. The whole population could be addressed.

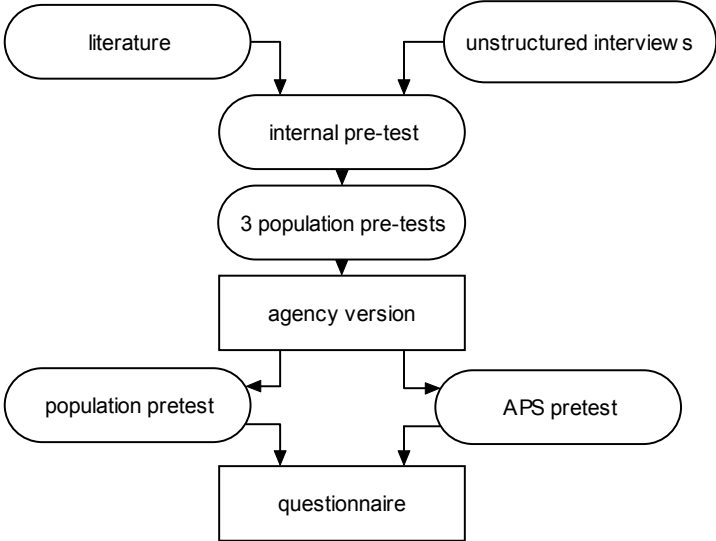


Figure 18: Development process of the questionnaire

Figure 18 represents the development process of the questionnaire. The foundation was the unstructured interviews in combination with the literature study (supra). We drafted a first version of the questionnaire based on the interviews and the performance measurement literature. Pre-testing was done in two phases (Robbins 1999). First, we discussed the survey with colleagues within the Public Management Institute. We should also mention that we sent out a similar survey to the Flemish autonomous agencies before we sent out the survey to the sections of the Ministry of the Flemish Community (reported in Van Dooren and Bouckaert 2004). The agency survey was pre-tested in a face-to-face context in three agencies. After each pre-test, we modified the survey. The experiences with the agency survey were an important input for the development of the survey of the Ministry of the Flemish Community. The version for the Flemish community was pre-tested in a section and in the Administration for Planning and Statistics. The latter organization has an overview position. This finally resulted in a questionnaire, which is appended in annexes (13.2).

The questionnaire has four subdivisions. First, we asked for an identification of the organization. The survey was not anonymous. We needed the name of the section because we intended to select cases for the study of the effects of performance measurement based on the survey. The non-anonymous nature of the survey is believed not to be problematic, since many questions deal with factual manifestations of performance measurement, i.e. in annual reports, budget documents, etcetera. Moreover, several questions were duplicated in the semi structured interviews which allowed for double checking the answers (see chapter 6). The identification also includes some questions on the characteristics of the section (size, budget, etcetera). The identification of the measurement practices in the section is the second subdivision. We strongly fall back on concrete documents. Asking for the measurement for performance in general would be too confusing. A typical wording of a question is “do you have indicators in you annual report?” By this, we induce a mental picture of the subject of the survey. In addition, assessments about the hindrances for performance measurement, the coverage rate, and the measurability of the output were included. The third subdivision analyzes the production of performance measurement. We asked, amongst others, for the actors involved, the data sources, data analysis and political involvement. The fourth section deals with the use of performance information. Again, we asked for the actors that are consuming information of the section. There were also questions about the political impact, the media attention and the internal management use.

## **5.4. Semi-structured interviews**

### **5.4.1. Definition**

As the name suggests, semi-structured interviews stand between structured and unstructured interviews. The interview framework is to a large extent a priori established. The interviewer must adhere to that framework. Yet, unlike structured interviews, there is some leeway in answering the

questions. Semi structured interviews will for instance not make use of multiple choice questions. The respondent needs to be able to provide context.

#### 5.4.2. Purpose

Semi-structured interviews aspire to combine the best of both worlds. They purpose the richness of the data that is acquired through unstructured interviewing as well as the comparability and systematic nature of the structured interviews and questionnaires. The structure of the interview allows for easier coding and analysis. The direction of the interview is more clear for both interviewee and interviewer. Both sides may be more at ease. This is not a trivial issue. In our study, respondents usually asked explicitly for receiving the questions of the interview in advance. When several interviewers are involved, the use of more structure in the interview reduces the possibility of interviewer bias.

Semi structured interviews have limitations too. First, they still may be too inflexible. The structure may be a straightjacket that gives the interview a unnatural character. Thoughts that are provoked during the interview may not fit into the predetermined framework, or only at a later stage. When the survey has a high-educated, professional population, as in our case, the researcher can make clear the internal logic of the interview and the objectives of the questions. This may to some extent remedy the inflexibility problem of the semi-structured interview. Sending out the questions for the interview in advance may also be instrumental in this respect. Yet, it should not be assumed that the interviewee has gone through the questions. Secondly, semi-structured interviews are face-to-face interviews. Therefore they are costly and/or time consuming. In this respect, they are similar to open-ended interviews or researcher administered structured interviews.

#### 5.4.3. Case selection and instruments in this study

The cases for the semi-structured interviews are selected based on the survey data. More details about the case selection and the instruments are provided in chapter 9.

### 5.5. Relation of data sources and research issues in this study

The methods for data collection and the research issues relate to each other in the way that is outlined Figure 19. The unstructured interviews were the starting point for the research project. These interviews were useful in for exploring the design parameters of measurement systems that are required for different uses of performance information (chapter 8). In addition, we formulated the research question on political demand and administrative supply from the interviews. On the one hand several interviewees did mention the issue of lacking political interest in performance information. Reports with performance indicators are said to end up in the drawer of decision-makers or – when

well edited – on the bookshelf. On the other hand, respondents mention the issue of having to answer to parliamentary questions. This seemingly paradoxical phenomenon led to the study of parliamentary questions and their measurement content (chapter 7). Yet, as we argued above, we needed additional document analysis for this study.

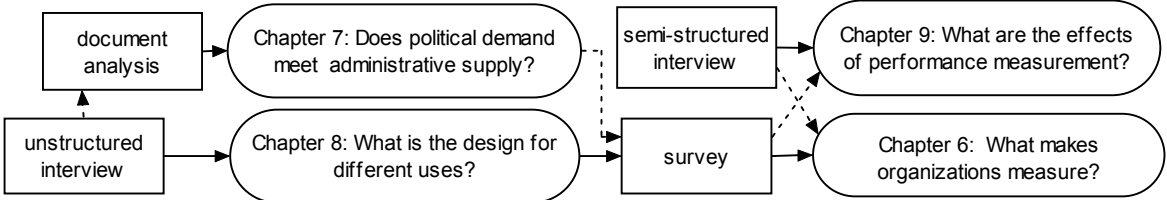


Figure 19: Relation of data sources and research questions

The two remaining research issues, why measuring performance (chapter 6) and the effects of measurement (chapter 9), are mainly explanatory in nature. The first research issue is studied by means of the survey. We will test six hypotheses on the use of performance information. We will make a distinction between adoption (“having performance measurement”) and implementation (“doing” performance measurement). The questionnaire will in second order also be used for selecting the cases for chapter 9. The effects of performance measurement are studied by means of semi-structured interviews. We will create a user profile and an effect profile and look for associations. We opted for semi structured interviews for countering social desirable answering. In a face-to-face interview, this is less likely. In second order, we used the interviews to check the reality of the hypotheses we formulated in chapter 6.





## 6. What makes organizations measure?<sup>53</sup>

### SUMMARY OUTLINE

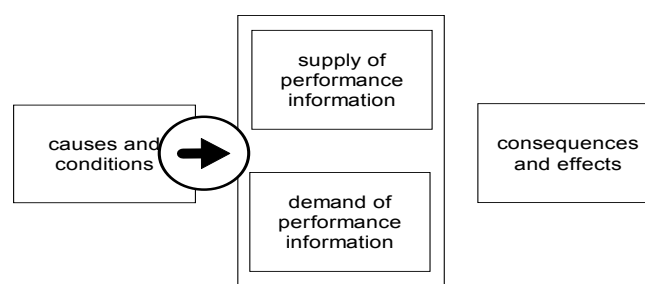
**Issue:** Although performance measurement is the Achilles heel of many public sector reforms, there is not often an explicit policy to obtain performance measurement information (Bouckaert and Peters 2002). Availability of performance measurement is assumed, and reforms are built on this weak foundation. Insight in the organizational and contextual factors that facilitate or impede performance measurement is crucial for developing a performance measurement policy. Governments may be tempted to design a one-size-fits-all policy, often based on the best practices in the public sector. However, differences between organizations may be considerable and should be taken into account. This section attempts to establish empirically some of the factors that explain differences between organizations.

**Research questions:** Six issues will be studied.

1. The measurability issues; is the type of output a determinant?
2. The political interest issue; does political interest or political apathy lead to measurement?
3. The scale issue; do large organizations measure more than big ones?
4. The street level discretion issue; does street level discretion play a role?
5. The means issue; is it mainly a matter of resources?
6. The goal orientation – issue; is the coupling with goals of main importance?

**Methodology:** survey of sections within the Flemish regional government (N=155, rr=48.3%)

### Graphical representation



<sup>53</sup> A previous version of this chapter is published in *Financial Accountability and Management: Van Dooren W. (2005) What makes organizations measure? Hypotheses on the causes and conditions of performance measurement.* *Financial Accountability and Management*. 12:3 363-383.

## **6.1. Problem definition and purpose**

A good insight into the organizational characteristics that foster performance measurement is crucial in order to set up a performance measurement policy. Therefore, insight into the factors that make organizations measure is needed. In this study, performance measurement is the dependent variable. Six issues on why organizations measure are studied based on the survey within the Ministry of the Flemish Community. Organizational characteristics are related to the adoption and implementation of performance measurement. A performance measurement policy is needed before a performance management policy may be set up.

Until now, the development of performance measurement has not been formally made compulsory for sections of the Ministry of the Flemish Community. Nonetheless, significant measurement initiatives are developed. This makes the sections of the Ministry of the Flemish Community an interesting population for research.

The chapter is structured along the variables and the hypotheses. Theoretical propositions therefore are embedded in the discussion of the hypotheses. The first paragraph shows how performance measurement is dealt with in this chapter. Next, we briefly describe the methodology of the survey. Thirdly, the dependent variables are presented. The dependent variables are performance measurement adoption and performance measurement implementation. Fourthly, by bringing in the independent variables six hypotheses may be made.

## **6.2. The role of information in organizations**

Before we describe adoption and implementation of performance measurement, we briefly describe the role of information in organizations, which is rooted in the information processing tradition of public policy research (Bobrow and Dryzek 1989). (Performance) information is seen as one factor in decision-making, amongst others. Lindblom and Cohen (1979) assert that information and analysis constitute only one route among several in social problem solving. Besides professional information, ordinary knowledge as well as daily interaction influences decision-making. Janis (1989) distinguishes cognitive decision rules, but also affiliative and self-serving decision rules. In our study, we follow this line of reasoning. Reliance on performance information is one approach to decision-making. The assessment of performance is based on several sources: both formal and informal (see also Ter Bogt 2004). Therefore, performance information should be *'fit for use'*. However, it certainly does not exclude other sources of information. This is not a purely positivist approach, nor a purely interpretative approach (Bobrow and Dryzek 1989). The underlying assumption is Simon's bounded rationality (Simon 1976). Since policy makers and top manager's rationality is limited by several factors, performance information has to have the right qualities for a specific purpose if it is to play a significant role.

### Independent variables

### Dependent variables

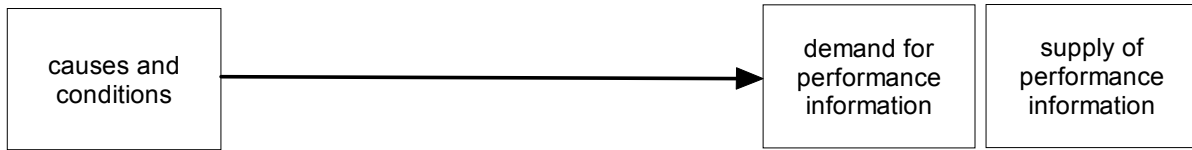


Figure 20: Outline of the study

The general outline of the study is represented schematically in Figure 20. The production process of performance measurement is the supply side. The uses form the demand/consumption side. A description of performance measurement in an organization may look at both the production process and the use of performance information. In this chapter, we will use the constituent parts of performance measurement in organizations to describe the dependent variable. The way organizations measure and use their information will be the starting point for composing two indexes; one for adoption and one for implementation of performance measurement and its uses (infra). The next step will be to look for the causes and conditions for adoption and implementation

### 6.3. Methodology

This study is based on a mailed survey of sections of the Ministry of the Flemish Community in Belgium. The Ministry of the Flemish Community is the regional administration of Flanders. It consists of seven departments in a matrix structure. The names of the departments provide some insight into the competences of the regional government in Belgium. Two departments administer staff functions such as budgeting, accounting, infrastructure, Human Resources and personnel services. The five functional departments are (1) Welfare, Health and Culture, (2) Education (3) Environment and Infrastructure (4) Science Innovation and Media, and (5) Economy, Employment, Internal Affairs and Agriculture. Each of the departments has several administrations. The third and lowest level is the section level.

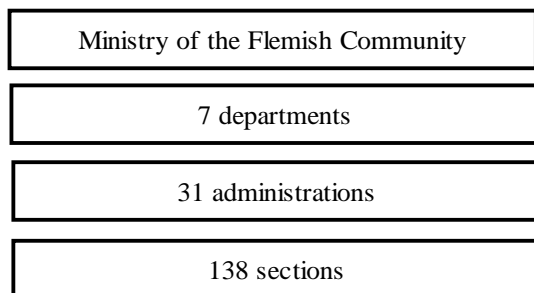


Figure 21: Organizational chart of the Ministry of the Flemish Community, Belgium

The survey was tested and modified in six subsequent face-to-face interviews. During the interview, the respondents could comment freely on the questions. They were encouraged to make their interpretations of the questions explicit. After every interview, the comments were incorporated in the survey. After the first interview, substantive changes have been made. The changes after the fifth interview were minimal.

The survey was divided in four sections (see appendix 13.2). The first section identified the organization. Items such as the number of full time equivalents, the budget and the policy sector were included here. The second section was about measurement practices. To what extent do organizations measure? In order to increase comparability over organizations, we referred to documents that all organizations understand in more or less the same way such as annual reports, budget documents, parliamentary questions, and quality models. The third section referred to the supply side. Who is measuring? Where does the come from? How are targets set? Is the information freely available? The fourth section was about the use of performance information. Who is using our information? What is the political involvement? What role does the information play in the internal management of the organization?

In total, there are 138 'afdeling' (sections). Yet there are several other subdivisions at more or less the same level as the sections. These are typically called 'cel' (cell), 'dienst' (service) or 'entiteit' (entity). Sometimes, they resemble the 'afdelingen'. In a formal sense, they are not 'afdelingen', yet in administrative reality they are. Sometimes, they are hierarchically subordinate to an 'afdeling' or they are labels for organizing the staff of the directorate-general. We sent the survey to 173 respondents; 138 sections are called 'afdelingen' and 35 would be 'afdelingen'. Of the 35, we kept 17 in the analysis. The rest was left out because they were or subdivisions of a section, or because they did not have a substantial saying on how to organize their unit. This was in particular the case for the 'cellen' (cells). Of the 155 remaining sections, 48.3% responded to the survey. There is a trade-off between response rate and length of surveys. According to Babbie (1975). This response rate is acceptable, given the length of the survey (approximately 200 items). It should be noted that there is no sampling involved. The whole population was included. It is reasonable to expect that organizations that measure more will be more inclined to respond to the survey. Berman (2000) made the same observation in the USA. Yet, extrapolation of the findings to the population of sections of the Ministry of the Flemish Community is not the purpose of this study. The issue is why organizations are measuring. The responses may be seen as a stratified sample of the population with an overweighing of users of performance measurement. Appendix 13.2 provides some descriptive statistics. There is sufficient distribution in the answers in order to study the factors that impact measurement.

The respondents could also return a hard copy of the survey or fill out a form on the Internet. More or less half of the respondents chose to use the version on the Internet, while the other half did send the hard copy back.

We repeated the six issues in the semi-structured interviews. The main purpose of these interviews was to study the effects of performance measurement (chapter 9). For that purpose, the case selection was deliberately skewed towards sections with a moderate to high use of performance measurement. The interview data was used to confront the quantitative analysis with more rich data. Although the main analysis thus is quantitative, we used some qualitative data to check the validity of the quantitative results. These qualitative evaluations are brief, and included throughout the text.

## **6.4. Dependent variables**

The overall dependent variable is the performance measurement system of the organization. The performance measurement system consists of the different uses (demand side) on the one hand, and the production process on the other hand. The aim is to classify organizations based on the degree to which they are measuring, i.e. the extent to which a performance measurement system is in place.

We approach the measurement of performance measurement in two ways; measurement of adoption and measurement of implementation. The distinction between adoption and implementation was made by Beyer and Trice (1982) to assess the utilization of social research. Recently, DeLancer Julnes and Holzer (2001) applied it on performance measurement in the United States of America.

Adoption and implementation in this chapter are not identical to production (supply) and use (demand) of performance information. DeLancer Julnes and Holzer (2001) borrow Stehr's (1992) notion of adoption as the development of a capacity to act. Implementation represents knowledge converted into action. For our purposes, adoption is about having performance measurement while implementation is about doing performance measurement. An organization in this case may have a broad range of indicators that are integrated in many policy and management tools, without doing something with it. That is, without relying on the information for actions such as changing strategies or operations, reallocating resources and evaluating personnel. Supply and demand thus refers to the measurement process and the tools for which the information is needed. Adoption indicates the extent to which an organization has the tools. Implementation refers to the extent to which the organization is doing something with the policy and management tools.

### **6.4.1. Adoption of performance measurement**

Adoption includes the set of behaviors through which decision makers decide that performance measurement may be useful for their organization (Beyer and Trice 1982). The outcome of the adoption process is that organizations have performance information. Behaviors in the adoption are about sensing and searching, affective reactions, selection and the (formal) adoption. These specific behaviors relate to more general organizational processes and components of behavior. Table 1 applies the model by Beyer and Trice on performance measurement for the adoption phase.

components of behavior	organizational processes	specific behaviors involved in the utilization process	applied on performance measurement
cognitions	information processing	sensing and searching	formulating questions: ✓ what is performance measurement? ✓ what are the costs and benefits? ✓ what are the alternatives?
feelings	affective bonding	affective reactions	attitudes: ✓ the attitudes towards information: hard data or soft and tacit knowledge ✓ a culture focused on results or on processes
choices actions	strategy formulation and control action generation	selection (formal) adoption	✓ integrating cognitions and feelings ✓ what are the indications of formal adoption of performance measurement?

Table 9: The adoption of performance measurement (based on Beyer and Trice 1982)

The measurement of the adoption of performance measurement looks at the last line in the table. What indications point to the acceptance of performance measurement and its uses in an organization?

The first approach to assess the degree of performance measurement adoption is the coverage rate. The Dutch Audit office used the coverage rate by indicators of the articles in the budget to assess the degree of measurement (Algemene Rekenkamer 1997). However, only the articles that could be explained in a meaningful way by using indicators were comprised in the assessment. According to the Audit Office (Algemene Rekenkamer) this was the case when it is technically feasible to determine a price and a homogenous quantity of public service provision. The activities that could not be measured were left out. In our survey, we asked the respondents to assess the coverage rate of the activities by indicators on a four-point scale from almost no coverage to an almost complete coverage.

A second way of looking at adoption is to look at formal documents such as the annual report (Boyne and Law 1991; Hyndman and Anderson 1995; Johnsen 1999a). In essence, it is an instrument for external communication to stakeholders. The inclusion of indicators in the annual report shows the adoption of performance measurement. An important advantage of using annual reports as a unit of analysis is its comparability. Across organizations, the annual report is well understood as a yearly reporting of the activities of the organization. Several authors analyzed annual reports as a proxy of the extent of performance measurement.

Thirdly, the quality movement often stresses the importance of measurement. Indeed, quality models such as the Balanced Scorecard, the model of the European Foundation of Quality Management, the Common Assessment Framework, etc. propagate performance measurement. Therefore, the use of quality models may be another indication of the adoption of performance measurement.

Fourthly, organizations may adopt performance measurement systems because it is an obligation. Obligations are often comprised in legislation. Our data however originated from a survey of section heads within the Ministry of the Flemish Community. The sections have two higher hierarchical levels, i.e. the director-general (DG) and the secretary-general (SG). We asked in the survey whether the section heads felt that performance measurement is an obligation of the DG or the SG, a recommendation of the DG or the SG, or neither one of them.

### 6.4.2. Implementation of performance measurement

Implementation includes the set of behaviors through which organizations actually carry out performance measurement. While the result of the adoption process is having performance measurement tools, the result of the implementation process is doing performance measurement. The capacity to act is put to work. Specific behaviors involved in utilization processes are diffusion, receptivity, commitment, evaluation, feedback, use and institutionalization (Breyer and Trice 1982). Table 2 applies these concepts on performance measurement

components of behavior	organizational processes	specific behaviors involved in the utilization process	applied on performance measurement
cognitions	information processing	diffusion	✓ the measurement initiative is made known in the organization: e.g. through training sessions
feelings	affective bonding	receptivity, commitment	✓ the organization is amenable to measurement
choices	strategy formulation and control	evaluation, feedback	✓ evaluation of the measurement system ✓ adjusting the measurement system
actions	action generation	use, institutionalization	✓ use of performance information ✓ performance measurement is more taken for granted

Table 10: The implementation of performance measurement (based on Beyer and Trice 1982)

Implementation goes beyond the formal structure. Indicators in documents such as the annual report point to a formal adoption of performance measurement. However, they do not give insight into the extent to which performance measurement and its uses in policy and management tools are put to work. Thus, indications of implementation are needed.

Firstly, performance measurement systems that are in operation are embedded in decision structure of the organization. In this case, performance measurement supports the decisions that an organization has to take. In most performance measurement textbooks, this is the prescribed use (Hatry 1999; Poister 2003). The purpose of measurement is to account for exceptional performance. These accounts finally may result in decisions to change the way in which things are done in the

organization. In our survey, we asked to what extent the performance information is used for the allocation of resources on the one hand and changing processes on the other hand.

Secondly, performance measurement system may be integrated in the actions of the organization. The actions refer to the day-to-day management and operations of the organization. Organizations that use performance measurement in their actions will have frequent internal meetings to discuss the measurement results. These meetings will not only be attended by top-management. Middle management and operators will also have to join the meetings.

### 6.4.3. Calculating adoption and implementation

We calculated two combined measures, i.e. one for adoption and one for implementation. The measure for adoption includes the coverage rate, the inclusion of indicators in the annual reports, the application of quality models, and the extent to which the adoption of performance measurement is an obligation of the higher hierarchical levels in the organization. The measure for implementation includes the frequency of internal discussion of results, the attendance by top and middle management of these sessions, the extent to which the performance information is used to change processes and the extent to which the performance information is used to take allocation decisions. There are four items for each of the two measures. Since all variables were ordinal, each item could be standardized by taking the average of the division of the observed values with the maximum value. Missing values were imputed in a 'pessimistic' way. We assumed that when respondents did not fill out an item, they would rather not measure. Thus, we for instance impute zero in a missing value on the question whether indicators figure in annual reports. We assume that if respondents do not mention it, they probably would not have it. Statisticians call this method cold deck imputation. Each missing observation is replaced with a pre-determined constant. Imputing the same value for all missing observations is justified if there is a strong justification for the choice of constant. Each standardized item has the same weight in the composite measures, i.e. 0.25. Table 11 represents the calculation schematically. Appendix 13.2 has the frequency tables for the items included in this chapter.



<b>Adoption</b>		<b>Scoring</b>	<b>Computation</b>
What is the coverage rate of the activities?		CR (0/1/2/3)	<p style="text-align: center;">ADOPTION =  <math>(CR/3 + AR/1 + QM/1 + (OB1 + OB2)/2) * 0.25</math></p>
Do indicators figure in the annual report of the section?		AR (0/1)	
Does your section apply quality models?		QM (0/1)	
Is performance measurement an obligation of higher hierarchy?	Obligation of Secretariat General	OB1 (0/1)	
	Obligation of Directorate General	OB2 (0/1)	
<b>Implementation</b>			
How often does your section discuss the results?		FR (1→7)	<p style="text-align: center;">IMPLEMENTATION =  <math>(FR/7 + (AT1 + AT2)/2 + PR/5 + AL/5) * 0.25</math></p>
What is the attendance by top and middle management?	Attendance of top management	AT1 (0/1)	
	Attendance of middle management	AT2 (0/1)	
To what extent is the performance information used to change processes		PR (0→5)	
To what extent is the performance information used to make allocation decisions?		AL (0→5)	

Table 11: Calculating adoption and implementation from the survey data

#### 6.4.4. Relating adoption and implementation

Organizations can be high or low on both implementation and adoption. The scatterplot shows the results of the combination of both scales.

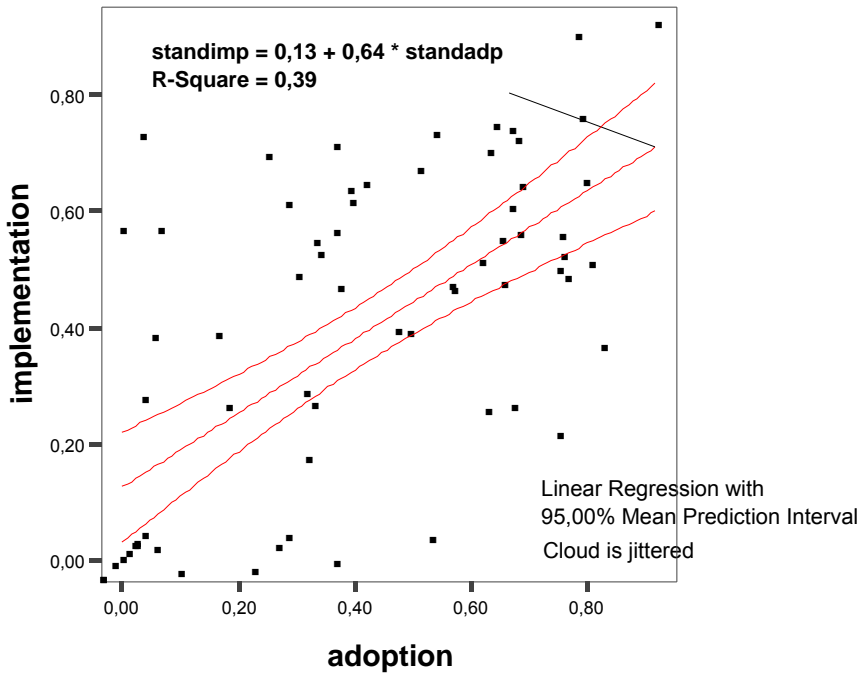


Figure 22: Adoption and implementation of performance measurement.

The correlation between adoption and implementation is significant<sup>54</sup>. The regression shows that adoption accounts for approximately 40% of the implementation. Some organizations do not adopt, nor implement performance measurement. Other organizations have high scores on both adoption and implementation. However, the scatter is widely dispersed. Many organizations have an atypical profile. Some organizations score high on adoption and low on implementation. Measurement in these organizations is probably predominantly outward oriented. They measure because they have to or because they see communicative opportunities in presenting performance data. They measure because they want to have it rather than that they want to do it. These may be the window-dressing cases, which have many measurement initiatives on the shelf, but do not use it in operations. Some

---

<sup>54</sup> Kendall's Tau-b, 0.432, n=72

organizations score low on adoption and high on implementation. These organizations do not have many formal initiatives, but the ones that are in place are being used internally. In sum, there are four profiles.

		Adoption	
		Low	High
Implementation	Low	No PM	Outward PM
	High	Inward PM	Full PM

Table 12: Profiles of performance measurement (PM) implementation and adoption

In this study, we will use implementation and adoption as the dependent variables. We will try to identify independent variables that explain why adoption and implementation are high or low.

**6.5. Independent variables, hypotheses, and results**

**6.5.1. The measurability hypothesis**

A first important independent variable refers to the characteristics of the output and outcome of the organization. This relation may be termed the measurability hypothesis. The basic proposition of this hypothesis is that some outputs and effects are easier to measure by nature. More tangible outputs and effects can be measured in a more precise way compared to non-tangibles. Hackman and Oldman (1980) distinguish between tasks with high routine and low ambiguity, tasks with average routine and ambiguity and tasks with a high routine and ambiguity. The measurement base is respectively time, group and progress. Measurement becomes increasingly difficult when routine lowers and ambiguity rises. In organizations with a high ambiguity and a low routine, measurement of output is more difficult. Typical examples would be embassies and cultural institutions. On the other hand, public housing corporations are a typical example of a sector with tangible outputs, i.e. the provision of social housing.

Other authors also point to the importance of the concept (Bouckaert and Halachmi 1995). Downs (1967) finds that the ease with which the results of bureau actions can be perceived and their effectiveness evaluated, is a structural aspect of an organization (p43). Wilson (1989) proposes a useful typology of organizations based on this division. He makes a distinction between four types of organizations; production, procedural, craft and coping organizations, based on the observability of their output and outcome. Hypothetically, measurement will be easier in production organizations and more difficult in coping organizations. Mol (1988) makes a distinction between whether outputs are identifiable and whether activities are homogenous. He finds that a performance budget is only feasible when activities are homogenous and outputs are identifiable. If not, the budget should focus on the process (homogenous activities but no identifiable outputs), the tasks (identifiable outputs but no homogenous activities) or even the input (no homogenous activities and no identifiable outputs).

Hofstede (1981) suggests that management control systems will be able to rely on performance measurement when goals are unambiguous, when activities are repetitive, when outputs are measurable, and when the effects of management are known. Finally, Blankart (1987) links the limits of privatization to service characteristics. He distinguishes between three types of consumption technology. First, inspection goods (raw materials, stationary, ...) may be privatized easily because of the tangible and measurable quality. Secondly, experience goods (e.g. advice, debt collection, ...) are also candidates for privatization because the market has overcome the deficiency of quality uncertainty through the extrapolation of experiences and the accumulation of goodwill by clients. Thirdly, trust goods (courts, police, general public administration) are difficult to privatize, because they are hard to evaluate, even through experience.

In our study, we focused on the output. Outcomes would be too hard to grasp in a survey. The respondents were asked to list maximum three of the most important outputs of the organization and to assess its measurability on a five-point scale. The mean of the three assessments was used to calculate measurability.

Obviously, this is a perceptual assessment. One could argue that the causal direction is different. Those who measure more could be more inclined to conceive output as measurable. Although this possibility cannot be ruled out completely, we are inclined to believe that the assessment is reasonably objective. Respondents had to name the three most important activities, which allowed for assessing the respondent's judgment. By naming the output, the assessment becomes very concrete. It is not a general intuition. Furthermore, we repeated the question on measurability in the semi-structured face-to-face interviews and found that respondents made a balanced judgment on measurability. We found that respondents made explicit comparisons with other sections in motivating low or high measurability, or that they explicitly referred to different activities within the section. Measurability is clearly an issue put through previous thought.

*H1a Organizations that have more observable outputs will measure more. Since it is easier to obtain highly valid and reliable performance data, it may be expected that implementation and adoption are higher in organizations with a measurable output.*

*H1b Organizations that find it harder to measure output will be more inclined only to adopt performance measurement.*

(Kendall-tau - b)		adoption	implementation
mean measurability	Correlation Coefficient	0,239**	0,209*
	Sig. (2-tailed)	0,007	0,018
	N	71	71

Table 13: Measurability hypothesis

Ad H1a. Organizations that have more measurable output both adopt and implement more performance measurement. Consequently, the implementation of performance management techniques such as performance contracts between central departments and agencies has to take into account these differences in output.

Ad H1b. The correlation is slightly stronger for adoption than for implementation. This may suggest that measurability is especially important for outward oriented performance measurement, while for an internal focus measurability is not equally important. The higher pressure on outward oriented measurement systems may provide an explanation. More outward orientation implies that more eyes are watching. This requires robust data, which is more easily obtained from tangible activities. The internal focus on the other hand allows for more leeway in the interpretation of the strengths and weaknesses of the data. Therefore, measurability may not be an equally important issue for implementation, which is mainly an internal matter. However, the difference in correlation is not strong, so the above explanation remains hypothetical.

### 6.5.2. The political interest hypothesis

Organizations often regard political involvement as an important explanatory variable for public sector reform (Pollitt and Bouckaert 2004). The support from elected officials is also an important factor for the deployment of performance measurement. Wang and Berman (2001) for instance found that involvement of elected officials is a determinant of outcome measurement, but less for output measurement. The latter is more of an internal management concern while the former relates more closely to citizens interests. The survey dealt with county governments in the USA with more than 50 000 inhabitants. Broom (1995) also identified political critical success factors for performance budgeting in five states in the USA. Among others, the executive and legislative branches need to question agencies about performance and the executive and legislative leadership is crucial. Other authors make similar arguments (e.g. DeLancer Julnes and Holzer 2001).

Although political participation appears to be important, it remains an intricate issue. First, the availability of information often outweighs the limited capacity to deal with it. Downs (1967) describes the way in which politicians deal with the information provided by the budget. They are looking at the increment, rely heavily on specialization (with a large staff of specialists) and tend to focus on areas that arouse public interest. Presumably, they will handle performance information the same way. Secondly, provided they have the capacity to deal with the information, they have to have the will. Performance measurement does not make politicians win or lose an election. However, the data that the performance measurement system yields may show weak performance. For politicians, everything they measure can be used against them, so they might claim the right to remain silent and not to measure at all. Politicians may indeed have disincentives to collect data (Behn 2001).

Two dimensions of political involvement were included in the survey. The first item asked whether the lack of political involvement is a potential hindrance for measuring performance. Secondly, the involvement of politicians is measured. The first measurement looks at the perception of the importance of political involvement. The second asks for the actual involvement.

*H2a. Political involvement has a positive impact on performance measurement, in particular on adoption.*

*H2b. Organizations that perceive the lack of political involvement as an important hindrance to measurement will have less adoption and implementation.*

(Kendall-tau - b)		adoption	implementation
"lack of political involvement is a hindrance" (1-5 scale)	Correlation Coefficient	-0,24117	-0,308
	Sig. (2-tailed)	0,011*	0,001**
	N	67	67
"politicians are interested in PI" (1-5 scale)	Correlation Coefficient	0,066	0,078
	Sig. (2-tailed)	0,539	0,453
	N	53	54

Table 14: Political interest hypothesis

Ad2a Political interest in the performance information does correlate with nor adoption, nor implementation of performance measurement. This reflects the originally inward oriented nature of the development of performance measurement in the Flemish (and Belgian) administration. When there is measurement, it was usually a bottom-up initiative of public managers rather than a government wide policy.

Ad2b. However, organizations that see political involvement as a potential hindrance have significantly less adoption and implementation of performance measurement. Although political interest is neutral to the implementation and adoption of performance measurement, the respondents see the lack of political interest as a potential hindrance. The lack of political involvement is an argument for not adopting and implementing performance measurement. However, in current practices the degree of political involvement is not explaining performance measurement.

The semi-structured interviews confirm the observation that political involvement is only loosely coupled to the development of measurement systems. We asked the respondents to assess the political interest in both the activities of the section and the information the section collects. Table 15

provides a schematic overview, including whether they are high users (H) or moderate users (M) of performance information<sup>5556</sup>.

		<i>Political interest in activities of the section</i>	
		Low	High
<i>Political interest in information</i>	Low	H H M	H H H M M
	High		H M M M

Table 15: Political interest in the information and activities of twelve sections (M; Moderate, H; high)

### 6.5.3. The scale hypothesis

Another factor that may influence performance measurement is the size of the organization (Kimberly 1976). Hypothetically, there are some effects of scale involved in the application of performance measurement. Lee and Burns (2000) found that larger states were less likely to backslide in performance budgeting initiatives than smaller states. The construction of performance measurement systems has fixed costs, which are easier to defray by larger organizations. Therefore, larger organizations may have more extensive performance measurement systems. The infusion of performance indicators for different outlets will be easier when a study-staff function is available. This will particularly be the case for the inclusion of indicators in annual reports and quality models. Scale is also hypothesized to have a positive effect on implementation. An increase in size may lead to an increase in problems of coordination and communication (Blau 1968). Top managers in larger organizations have a higher span of control, are more impersonal and therefore will have a higher need to manage internally by numbers. Managers in smaller organization may rely on management that is more interpersonal.

*H3 Larger organizations will measure more. We expect scale to be an important factor for adoption as well as having a positive effect on implementation.*<sup>57</sup>

(Kendall-tau - b)		adoption	implementation
fte total	Correlation Coefficient	0,218**	0,191*
	Sig. (2-tailed)	0,009	0,020
	N	72	73

Table 16: Scale hypothesis

<sup>55</sup> The user profile is based on the qualitative assessment of the semi-structured interview. More details about profiling can be found in chapter 9.

<sup>56</sup> Note that the selection of the sections for the semi-structured interviews has a moderate to high profile. We did not select low users for the semi-structured interviews, since the main purpose of these interviews was to study the effects of performance measurement.

<sup>57</sup> The range of organizations in terms of FTE is represented in appendix 13.2.

Ad H3. Scale correlates positively with both adoption and implementation. Large organizations measure more than smaller organizations. This result is important - although not surprising. It raises questions about the required capacity for performance measurement. An often-heard complaint is the administrative pressure that performance measurement systems put on organizations. Small organizations may not be capable to deal with the requirements of putting up a performance measurement system. High ICT requirements in particular may prohibit performance measurement in small organizations. Therefore an important issue is the design of low-keyed performance measurement systems that can be implemented in small organizations.

The sections of the semi-structured interviews did not experience problems of scale. The scale was in general considered sufficient for their respective measurement efforts. Yet, it was the case that the sections with a high user profile are also the largest ones. The indications regarding problems of scale point to sections that are too small rather than too big. Several sections did mention that assistance at the departmental level is important for further development of measurement – in particular for ICT support. Three sections mentioned that they are probably too small for implementing a measurement system. One section acknowledged that the measurement effort is heavily depending on one or two employees. This affirms the need for studying the potential of low-keyed performance measurement.

#### 6.5.4. The street level discretion hypothesis

The freedom of operators to define their job is considerable in the public sector. This discretion leads up to a considerable decisive power. Lipsky described this phenomenon in street-level bureaucracies (Lipsky 1980). Wilson (1989) expands this argument to all bureaucracies and provides several explanations for this rank-and-file leeway. The goals of public sector organizations are often vague and ambiguous. The tasks of the operators often do not have a univocal purpose. Moreover, middle and top management are more concerned with the constraints of the public sector than with the day-to-day operations. The tasks will thus be defined in other ways. The situational imperatives - the concrete circumstances with which the operators are confronted - are an important factor. Police officers for example may cope with a situation rather than do it as per instructions. According to Wilson, the situational imperatives will be stronger when the services are face-to-face with clients, and when the clients are uncooperative.

Performance measurement systems may be introduced to monitor the behavior of the operators. However, these performance measurement systems are likely to encounter resistance from within the organization. Therefore Berry et al. (2000) assert that agencies whose street level workers command significant professional discretion will find performance measurement more difficult. It should be noted that discretion is to a certain extent determined by the nature of service delivery, and therefore there is a conceptual linkage with the measurability hypothesis.



In our study, the indication of street-level discretion is the existence of a system for time registration may point to lower discretion. Time registration usually assumes a standardization of operating procedures. The smaller the registration unit (15min, hour, day...), the less discretion. The scale starts from zero (no time registration) to four (registration unit  $\leq$  1 hour).

*H4. Low discretion will lead to a high implementation. The adoption may be high or low.*

<i>(Kendall-tau - b)</i>		<i>Adoption</i>	<i>Implementation</i>
Time registration systems	Correlation Coefficient	-0,030	0,286
	Sig. (2-tailed)	0,798	0,011(*)
	N	50	50

Table 17: Street-level discretion hypothesis

The existence of strict time registration systems is indifferent to performance measurement adoption. Organizations that operate time registration do not adopt performance measurement more. However, the linkage with implementation is (borderly) significant. Organizations that operate time registration systems have a higher degree of implementation. If time registration were a good proxy for discretion, this would imply that lower discretion indeed would lead to a higher implementation of performance measurement.

The evidence from the interviews is somewhat undecided. We noted several reactions. First, some section managers saw discretion as a good thing. They saw their role as enablers of initiatives. A strict interpretation of the law would be dysfunctional. Secondly, some sections explicitly reject the role of measurement systems in this regard. They rather rely on professionalism of the employees. Thirdly, several sections acknowledged the potential of measurement for reducing discretion. Usually, discretionary freedom aligned through the study of outliers.

**6.5.5. The means hypothesis**

The fifth hypothesis is about the impact of the lack of financial, human and ICT resources on performance measurement adoption and implementation. Previous research often pointed to insufficient resources as an explanation for implementation failure (i.e. having without doing). Ammons (1986) found that only very modest amounts of executive and staff time are devoted to the appraisal systems in cities in the USA. Berman and Wang (2000) refer to the importance of capable staff to come to 'high use' of performance measurement information. They add, among other things, adequate information systems as a requirement for measurement capacity. DeLancer Julnes and Holzer (2001) and Grizzle and Pettijohn (2002) identified the commitment of resources as one of the factors for performance measurement. Our survey asked for the extent to which the respondents see the lack of ICT, human and financial resources as a potential hindrance for performance measurement.

*H5. More performance measurement adoption and implementation is expected in organizations that do not perceive the lack of resources as a potential hindrance to measurement.*

(Kendall-tau - b)		adoption	implementation
lack of ICT resources (1 low - 5 high)	Correlation Coefficient	-0,010	-0,061
	Sig. (2-tailed)	0,912	0,497
	N	70	71
lack of financial resources (1 low - 5 high)	Correlation Coefficient	0,015	-0,051
	Sig. (2-tailed)	0,871	0,583
	N	66	67
lack of human resources (1 low - 5 high)	Correlation Coefficient	-0,028	-0,027
	Sig. (2-tailed)	0,760	0,763
	N	69	70

Table 18: Means hypothesis

Surprisingly, the perceived lack of resources does not explain the degree of adoption and implementation. However, the descriptive statistics of the perceived lack of resources reveal that lack of resources is an issue. The average on a five-point scale (1 low barrier - 5 high barrier) is 3 for ICT (Information and Communication Technology) and 4 for human resources (figure 3). The issue of mobilizing sufficient means to adopt and implement performance measurement does not disappear when a measurement system is in place. In other words, all organizations feel that they do not have enough resources to measure. Some cope with it while others do not.

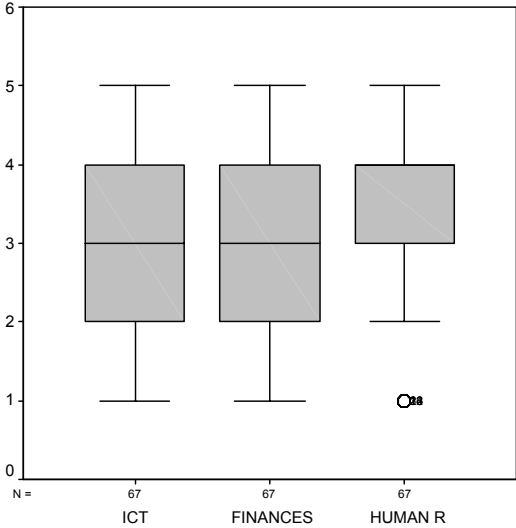


Figure 23: Lack of resources as a potential hindrance for measurement.

The semi-structured interviews seem to confirm this picture. Within the moderate and high users that are part of the case studies, some experience a lack of resources and others don not. Three out of six

of the high users and five out of six of the moderate users mention a lack of resources. The most pressing problem seems to be personnel, although in these cases, ICT problems are prevelant.

### 6.5.6. The goal orientation - hypothesis

Finally, the goal orientation may explain performance measurement. The connection of performance measurement with the goal setting processes of the organization is often considered crucial (Moynihan and Ingraham 2003). The connection with strategic planning for instance is important for performance budgeting (Berry, et al. 2000). Indicators in annual reports of district councils in the UK fail to state explicit targets and goals (Boyne and Law 1991). Indicators that are decoupled from the goals risk irrelevance (Mol 1996). However, a multiple case study by Johnsen (1999b) did not support the hypothesis. He found that the successful cases used a decoupled implementation mode, i.e. by not linking performance indicators to objectives. Decoupling may lead to an increased instrumental use of performance information as opposed to the symbolic and rhetorical use of performance information. Our survey asked for the extent to which the respondents see the decoupling of indicators and goals of the organization as a potential hindrance for performance measurement.

*H6. Decoupling of goals of the organization and the indicators will lead to less measurement. This will especially be the case for adoption, and less for implementation.*

(Kendall-tau - b)		adoption	implementation
HIND_GL	Correlation Coefficient	-0,114	-0,189
	Sig. (2-tailed)	0,233	0,046*
	N	66	67

Table 19: Goal orientation hypothesis

Ad H6. Organizations that consider the decoupling of performance indicators from the organizational objectives to be a potential hindrance for performance measurement development have significantly less implementation of performance measurement. Apparently, the linkage with objectives is important for decision-making and changing actual processes based on performance measurement. However, the relation is not significant for what adoption is concerned. Organizations that experience a potential decoupling of goals and indicators do not necessarily have less adoption of performance measurement. Decoupling is more troublesome when doing performance measurement (implementation) than for having performance measurement (adoption). The cases of the unstructured interview did not report problems of decoupling.

## 6.6. Conclusions

Performance measurement in this chapter was the dependent variable. It was measured in two ways; adoption and implementation of measurement. Adoption refers to formal acceptance of performance measurement and the policy and management tools. Implementation is about the integration of performance measures in the operations of the organization. Adoption and implementation are related. More adoption usually implies more implementation. However, other profiles exist. Some organizations have a high formal adoption, but low implementation and vice versa. The former have a more outward oriented measurement system. The latter are inward oriented.

Now what makes organizations measure? Six organizational factors have been studied. First, measurability of the services of the organizations is a key factor for implementation. Organizations that have more routine-based services have a higher implementation and adoption of performance measurement. Although this is not a novel observation, it has important practical consequences for the introduction of performance related policy and management tools. A successful implementation of these policy and management tools will require specific guidelines on how to measure performance in services that are not based on routine, but on human relations and ideas.

Secondly, political interest is in the Flemish case not an important factor for adoption and implementation. The degree of political interest for measurement does not explain the degree of adoption and implementation. However, organizations that see the lack of political interest as a potential hindrance, will measure less. There is a divergence between actual involvement and the potential impediment.

Thirdly, scale is also relevant. Large organizations measure more. This invokes questions about the minimal capacity that organizations need to measure. In the Flemish case, the best practices that serve as a role model for performance measurement are mostly large organizations. A performance measurement policy should also include guidelines on measuring performance in small organizations. Presumably, additional research is needed on this issue. What can (not) be done in small organizations?

Fourthly, street-level discretion -with the existence of time registration as an indicator - correlates positively with implementation but not with adoption. Less discretion (refined time registration) leads to higher implementation of measurement. Unpredictably, this is not the case for adoption.

Fifthly, the potential impediment of the lack of resources does explain neither implementation nor adoption. Organizations that measure as well as organizations that do not measure perceive the lack of resources as a potential barrier to performance measurement. In particular, human resources are

considered important. The provision of sufficient resources is a critical factor for starting up as well as maintaining and extending performance measurement.

Finally, the linkage between goals and indicators seems to be of particular importance for the implementation of performance measurement. The decoupling does not seem to impede adoption. However, when the final objective of the measurement initiatives is the implementation of performance measurement, the linkage between goals and indicators is important.

Insight into the organizational and contextual factors that facilitate or impede performance measurement is crucial for developing a performance measurement policy. Governments may be tempted to design a one-size-fits-all policy, often based on the best practices in the public sector. However, differences between organizations may be considerable and should be taken into account. This study attempted to establish empirically some of the factors that explain differences between organizations.



## 7. Does administrative supply meet political demand? A cross-sectoral comparison.<sup>58</sup>

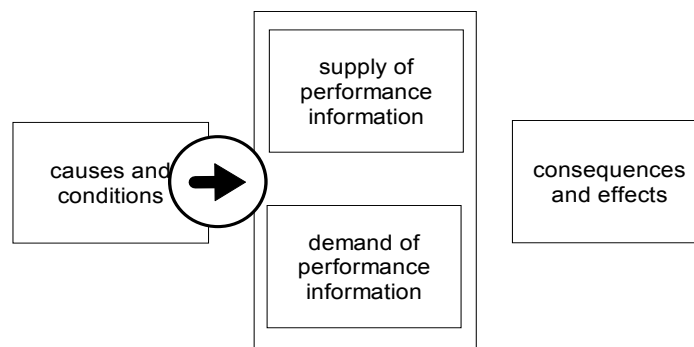
### SUMMARY OUTLINE

**Issue:** Civil servants often criticize politicians for not using performance information. Politicians complain that reliable performance information is not available. Politicians and civil servants complain that their own policy sector is hard to measure, if not immeasurable, compared to other sectors. This section seeks an empirical insight into the administrative supply and political demand of performance information and investigates whether there are differences between policy sectors.

**Research questions:** Three sets of research questions will be studied. First, to what extent is there a supply and demand of indicators? To what extent does supply and demand meet? What is the quantity of the (mis)match? Second, what is the quality of the supply? What are the motivations, if any, for not providing the demanded information? An additional assessment on quality is the explanation for not providing the demanded indicators. Thirdly, are the differences between policy sectors significant and if so, which policy sectors measure more?

**Methodology:** The methodology is a documentary analysis of parliamentary proceedings, i.e. the Members of Parliament's (MP) questions to the executive in the Flemish parliament (N=270)

### Graphical representation



<sup>58</sup> A previous version of this chapter is published in Public Management Review: Van Dooren W. (2004) "Supply and demand of policy indicators: a cross sectoral comparison" *Public management review*. 6 (4): 511-531.

## 7.1. Introduction

Public organizations in Flanders have developed datasets that have a considerable potential for evidence-based policies. However, how much of this information reaches the political arena? On the one hand, the lack of political interest in policy indicators is a frequent complaint of civil servants (supra). On the other hand, politicians and political cabinets touch on the lack of data availability. In their view, data often does not exist, is not timely or is not accurate.

The accurate crime statistics is one example of the administrative failure to provide incontestable data. Measurement in the police force is an increasingly important issue in Flanders (Van Dooren 2004). However, the failure to provide accurate crime statistics highlights the mismatch between supply, which is of a too low quality, and the demand, which may be of a too high aspiration. Ponsaers and Bruggeman (2005) conclude that government simply cannot tell whether crime went up or down. The crime statistics have breaks in time series. Moreover, aggregation is problematic due to differences in definitions.

Situations like these may be worrisome from both the civil servant's and the politician's perspective. The civil service complains that their data is not used. Politicians complain that the civil service does not have the right data, at the right time and of the right quality. The challenge is to work for a better match between demand and supply. This section first briefly states the research questions. Next, we point to the importance of the policy sector as a level of analysis. Then, we describe the data and the methodology. As pointed out above, it is an analysis of the content of parliamentary questions. Next, we present the results and draw the conclusions.

## 7.2. Research Questions

Three sets of research questions are studied.

- First, to what extent is there a supply and demand of indicators? To what extent does supply and demand meet? What is the quantity of the (mis)match?
- Second, what is the quality of the supply? What are the motivations, if any, for not providing the demanded information? We also assess the explanation as to why the administration did not provide the demanded indicators.
- Thirdly, are the differences between policy sectors significant and if so, which policy sectors measure more? Are the differences mainly input, output or outcome indicators?



### 7.3. Policy sector as a unit of analysis and a meso level reality

Several academics stress the importance of policy sectors as a unit of analysis. According to Benson, policy sectors comprise “a *‘cluster’ or complex of organizations connected to each other by resource dependency relationships and distinguished from other clusters or complexes by breaks in the structure of resource dependencies*” (Benson 1982: 148). Analysis should, in his view, focus on the policy sectors. Three levels may be examined. First, there is the administrative structure, which is the network of agencies in resource dependent relationships. A deeper level of analysis is the interest structure, which represents the demand, support, administrative provider and coordinating groups with interests in the policy sector. Thirdly, rules of structure formation limit and enable action at the two previous levels. These are the *‘deep rules’* that admit and exclude demands from the policy making process and that limit the choices and behavior of different parties in the policy sector. The supply and demand of indicators, or lack of them, may be seen as a method of structure formation. By choosing indicators, parties might also select issues and construct demands.

Sabatier’s theory on advocacy coalitions digs deeper into the relations within this cluster of organizations. The most significant point to stress here is his acknowledgement of the importance of policy sectors (1999). One of his basic premises is “*that the most useful unit of analysis for understanding policy change in modern industrial societies is not any specific governmental organization or program, but a policy subsystem (or domain).*” Within these subsystems, there are usually one to four advocacy coalitions, consisting of administrative agencies, legislative committees, interest groups, policy analysts, researchers, journalists and other levels of government. Advocacy coalitions will adopt strategies into further their policy beliefs; e.g. changes in rules, budgets and personnel. The demand and/or the supply of indicators could be seen as a way of translating the policy beliefs.

The extent to which policy sectors rely on indicators may differ between policy sectors. In some policy sectors, a quantitative translation (i.e. with indicators) of policy beliefs will prevail, while other sectors will rely more on a qualitative description. Sector-bound institutional factors may influence the magnitude of indicator-based policies. DiMaggio and Powell (1983) predict that in conditions of uncertainty (which are common in the public sector), organizations will be copying or imitating other organizations that are seen to be successful. This process is termed mimetic isomorphism. Following the aforementioned theories, considerable differences between policy sectors are expected in both the content and quantity of measurement.

## 7.4. Data and Methodology

The study is based on a content analysis of parliamentary proceedings. A quantitative systematic content analysis has been carried out. Content analysis is a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding (Krippendorff 1980). Starting from a predetermined input/output model (infra), the parliamentary questions of MPs' and the answers of the members of the executive have been examined for the occurrence of indicators.

The MPs' questions constitute the political demand for indicators. It is a dependable proxy for the demand of indicators in a policy community. Often, interest groups and other third parties prompt the MPs' questions. Therefore, it is assumed that the occurrence of indicators in the MPs' questions indicates the use of indicators in the broader policy community. The answers of ministers on the other hand constitute the proxy of the administrative supply of indicators. Usually, the ministers pass the questions on to the administration. The administration then formulates an answer with which the minister returns to parliament. Therefore, the occurrence of indicators in the answer might serve as a useful proxy for the availability of indicators in the administration. The indicators in the parliamentary questions and answers are thus proxies for broader concepts, i.e. the demand for information in a policy sector and the supply of information by the administration. It is important to look upon the parliamentary questions as an approximation of supply and demand only. The measurement method used in this chapter is -as most measurement methods do - only representing that part of reality.

The following variables have been analyzed for each parliamentary question and the corresponding answer.

▪ V1 policy field:	the policy sector to which the question refers
▪ V2 demand (input)	the number of input indicators in the question
▪ V3 demand (intake)	the number of intake indicators in the question
▪ V4 demand (output)	the number of output indicators in the question
▪ V5 demand (effect)	the number of effect indicators in the question
▪ V6 demand (context)	the number of context indicators in the question
▪ V7-12	idem as V1-6 but for the answer that is provided (supply)
▪ V13 direct link:	how many indicators provide a precise answer to the indicators in the question?

Figure 24: Variables for the content analysis of parliamentary questions

The input-output-effect model needs some more elaboration. We presented the general framework above. Here we demonstrate how we applied the mode in the content analysis with special attention categorizations that are debatable. Although the input-output model may seem univocal, several

interpretations challenge a straightforward interpretation the model. We provide our interpretation. Yet, other interpretations are conceivable.

Input indicators refer to the resources that go into the public sector system. Mostly, these are put in financial terms. However, other production factors such as personnel and infrastructure may also be considered. Indicators on tariffs and user charges are also included in the input category, since tariffs are usually the complement of the investment of public means. In addition, the amount and number of subsidies to organizations that provide services to third parties on behalf of the government are seen as input indicators. Examples are subsidies to museums, schools, hospitals, etc. These organizations play an important role in the provision of public goods, but are formally independent from government. However, questions on the amount and the number of subsidies to individual citizens and companies are considered as output, since these actors do not belong to the public sector apparatus. In these cases, the subsidy serves as a policy instrument rather than as a way of financing the provision of public goods.

Intake indicators refer to the 'unprocessed material' with which the public sector works. Obviously, the 'unprocessed material' is only a metaphor. Mostly the intake consists of people, but it could also be forests, houses, etc. Note that both indicators on the actual intake and the potential intake are included in this category. In addition, often MP's ask for indicators about the people that are not taken in. These non-intake indicators are also included here.

Output indicators indicate how many goods and services the public sector provides. The public sector organizations have a high degree of control on this. These indicators point to what the public sector offers to society in terms of products and services. Occupancy rates are equally seen as output indicators since they give an indication of over- and undersupply of public goods. Finally, also subsidies to individuals and companies are seen as output indicators (supra).

The effect indicators comprise both the effects in the short and the longer term. Several authors made this distinction (Hatry 1999, Ammons 1995), which is particularly relevant for organizations that develop measurement systems. In the context of the aforementioned research questions (supra p.5) however, the main interest lies in the fact that both short and long term effect indicators deal with the consequences of public goods and services in society. Effect indicators assess the societal reaction to output. Waiting lists and dropout rates from programs are seen as short-term effect indicators, assessing the impact of a program. Of course, the output of an organization is important to explain dropout or waiting lists. However, the dropout of e.g. a training program and the registration on a waiting list are reactions of societal actors to public sector output.

Contextual indicators refer to factors that influence the effects of a policy although these contextual factors are not the subject of the policy. Economic growth is often a typical context indicator. Note that a given indicator may be a context indicator in one case and an effect indicator in another case. The unemployment rate in a region is clearly an effect indicator for an Employment Office, while it may be a context factor for cultural participation. It is important to distinguish contextual indicators from effect indicators, since the former are being used to explain why the latter did (not) occur. Effects lay within the government's reach, while the manipulability of contextual factors is considered low.

Finally, in order to assess the quality of the answers, it was analyzed whether the indicators in the answer were a precise answer to the question. A direct link between indicators in question and answer signifies a qualitative match between supply and demand because the exact data that has been asked for is provided. The absence of a direct link points to a supply of other indicators than the ones asked for.

The analysis comprises 270 parliamentary questions from nine policy sectors within the competence of the Flemish regional government. The parliamentary questions date from 2000-2003<sup>59</sup>. Thirty parliamentary questions per policy sector are selected at random. The policy sectors covered are described below. The division is based on the parliamentary commissions.

Frequency tables for the variables are appended (appendix 13.3)

---

<sup>59</sup> The questions and answered are periodically published in the bulletin of "questions and answers". Only the questions that are answered on time are included in the analysis in order to have a homogenous population. Questions have been drawn of the bulletins of mei/00, sep/00, okt/00, nov/00, jan/01, feb/01, mrt/01, mei/01, aug/01, okt/01, dec/01, jan/02, apr/02, jun/02, jul/02, aug/02, sep/02, okt/02, dec/02, and jan/03.

- Education; all educational matters except for the age of compulsory attendance
- Internal affairs; legislation and organization of lower tiers of government (however, the local police and fire departments are a federal competence)
- Culture: the arts, socio-cultural affairs, libraries, etc
- Foreign affairs; foreign trade, development, some multi-lateral initiatives (however, the diplomatic corps comes under the federal government)
- Environment: environment, nature, land and water management
- Mobility and public works: roads, traffic, waterways, and ports
- Employment: job training and employment finding (however, the payment of unemployment benefits and inspection on moonlighting are federal competences)
- Welfare and health: policies for disabled persons, the elderly, youth, hospital management, etc. (however, all aspects relating to social security are a federal competence)
- Housing: social housing, supply and demand on the market

Figure 25: Policy sectors for the analysis of MP's questions.

Three examples of parliamentary questions are included.

#### Examples of the analysis of the Parliamentary Questions

Q1 MAINTENANCE OF THE WATERWAYS - 29 NOVEMBER 2002

**Question:** What has been spent on dredging operations since 1996 (input indicator)? How many vessels have been employed (input indicator) and how much silt has been cleared away (output indicator)? Have any incidents been reported that are related to an insufficient draught (effect indicator)?

**Answer:** The amounts spent, the vessels employed (input indicator with a direct link) and the quantities of dredged silt (output indicator with a direct link) are listed for each basin. Furthermore, it is answered that no incidents related to an insufficient draught are reported (effect indicator with a direct link).

**Total for this observation:** 2 input (demand), 1 output (demand), 1effect (demand), 2 input (supply), 1 output (supply), 1effect (supply), 4 direct links. This is an observation with high quality and quantity information.

Figure 26: An example of the content analysis of parliamentary questions: maintenance of waterways

Q2: AMERICAN INVESTMENTS - STATUS QUESTIONES - 21 AUGUST 2002

**Question:** Do American investors avoid Flanders? What has been done to attract American investors? What initiatives have been taken to enhance the competitiveness of Flanders, particularly with regard to the USA?

**Answer:** The decline of foreign investments in recent years is a global phenomenon. In 2001, foreign investment in all industrialized countries fell 59% (context indicator). However, UNCTAD data shows that the share of Belgium and Luxembourg in the worldwide investment flows is 13.8 times its share in the world economy (effect indicator), which is the number one position worldwide. Moreover, investments are mainly in manufacturing and chemical industry (72% in Belgium versus 29% in the Netherlands and 32% in the UK) (effect indicator). These sectors are more than other sectors subject to delocalization.

**Total for this observation:** no indicators in the question. 1 context and 2 effect indicators in the answer. This is an example of a supply without a demand.

Figure 27: An example of the content analysis of parliamentary questions: American investments

Q3: INSURANCE FOR THE NEED OF CARE FOR HANDICAPPED PERSONS - 21 FEBRUARY 2002

**Question:** A scale with the severity of the handicap is used for determining whether a handicapped person is eligible for the insurance. What is the profile of the persons that are just not entitled to benefit from the insurance (intake indicator)?

**Answer:** More than 50 000 handicapped persons will benefit from the insurance (intake indicator). For the moment, there is no insight into the number of handicapped persons that score just below the eligibility line.

**Total for this observation:** 1 intake indicator in the question, 1 intake indicator in the answer, no direct link.

Figure 28: An example of the content analysis of parliamentary questions: need of care for the handicapped

## 7.5. Results

In total, 157 parliamentary questions (58%) did include indicators, either in the question, or in the answer. Indicators apparently play a role in politico-administrative arena. The power of expression of numbers is apparent. The MPs asked (52%) for indicators in 141 cases. There were on average 1.92 indicators for each question. The ministers and their administrations included indicators in 125 (46%) answers with an average of 1.99 indicators for each answer. It seems that demand is higher than supply (141 questions compared to 125 answers).

Figure 29 confirms. It represents a linear regression of supply and demand. The X-axis is the total number of indicators in the question (total demand) and the Y-axis is the total number of indicators in the answer (total supply). The regression line lies underneath the 45° line that would represent a perfect match between supply and demand. On average, supply is lower than demand. Roughly, the demand explains 40% of the variance in the supply. This implies that 60% of the variance is not accounted for. Other variables may come to the fore. In this chapter, we focus on the variable 'policy sector' is looked into in more detail.

## Supply and demand curve

indicators in the answer by indicators in the question

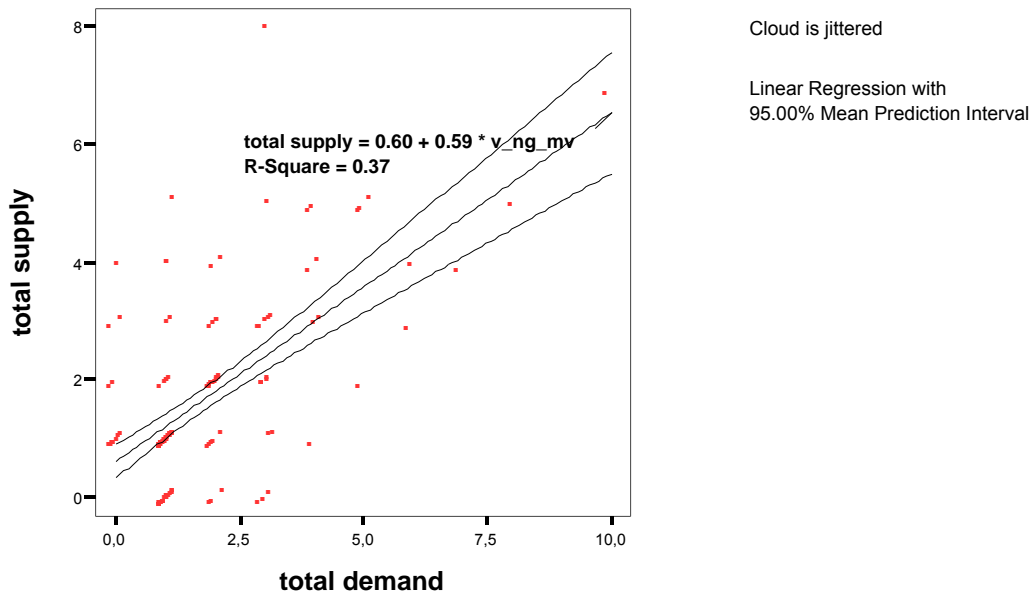


Figure 29: Supply and demand of indicators.

We also analyzed the quality of the answers. It was assessed whether the indicators in the answers related directly to the question. A 'direct link' between answer and supply entails a higher quality. 39% of the answers provided by the ministers and the administration to the MP's questions did not have a direct link to the question. The reasons are threefold. First, in 16 cases (10.2%) there were no indicators in the question. Nevertheless, the minister included indicators in the answer. In this case, there is no problem with quality as defined in this chapter (i.e. as a direct link between question and answer). There cannot be a link, since there are no indicators in the question to link to. The indicators in these cases are provided as an argument to a qualitative question. Secondly, in 16 (10.2%) cases the answer takes in other indicators than the ones asked for. This is the case when for instance exact data is not available. Other indicators are used to approximate the precise question. Thirdly, in 30 cases (19.1%) there were no indicators in the answer, although they had been explicitly asked for. Even when there is a direct link, the question remains whether it matches the demand completely or not. If for instance five indicators are demanded, and only one indicator is provided, then this parliamentary question would fall in the direct link box. However, supply does not match demand. In 70 cases (44,6%) there is an exact match. In 25 cases (15,9%), there is undersupply. Not all the requested indicators are provided.

Direct links	Frequency	Percent
No	62	39,5
<i>REASONS</i>		
1. no demand, but there is supply	16	10,2
2. demand, but other indicators in supply	16	10,2
3. demand, but no indicators in supply	30	19.1
Yes	95	60,5
1. exact match	70	44.6
2. demand > supply (direct link)	25	15.9
Total	157	100,0

Table 20: Frequencies of the occurrence of direct links (*In parliamentary questions with indicators*)

Often, no reason is given for not providing the requested indicator. In these cases, the questions usually are answered in a descriptive way. However, several times the absence of the demanded indicators is explicitly motivated. The explicit motivations for not providing an indicator fall into four categories:

- Availability. This is the most obvious category. The information cannot be provided because it is simply not available. Mostly, it is acknowledged that the administration did not make the necessary registrations for answering the questions. In some cases, however, it is questioned whether it is possible to collect the data at all. It was doubted for example whether the exact societal cost of a shortage in the Hepatitis B vaccine could be calculated. Often MP's entertain excessive expectations on the information system of the administration.

- Timeliness. Data is available, but it is not up to date. An example is the census data. The census is carried out every ten years. Therefore, the answers often had to refer to the census of 1991, which makes it difficult to formulate an accurate answer to a question.

- Coordination. Coordination problems are the third motivation for lacking data. Frequently, different agencies at different tiers of government have to cooperate in order to answer a parliamentary question. In a federalized country such as Belgium, these coordination problems are aggravated. For example, mobility is a regional competence. The regions are responsible for the organization of the roads. However, compliance with traffic rules is a federal competence and most of the data on the latter aspect is found there.



- Consolidation. The fourth explanation is the lack of consolidation. The data is available, but they are fragmented. The bits of information are scattered in different organizations. In the data set, consolidation problems surfaced for local authorities, schools, housing corporations, committees for special youth care and local payment bodies for the 'zorgverzekering' (insurance for care needs).

The total number of indicators and the count of parliamentary questions that include indicators in either the question or the demand show significant differences between policy sectors. Sectors that have the main indicator orientation are welfare and public health, housing and employment. Culture, education, internal affairs have notably fewer indicators. Foreign policy, mobility and public works and environment have a midpoint position.

policy sector	Parliamentary questions with indicators (freq)	total number of indicators (count)
<i>internal affairs</i>	10	43
<i>education</i>	13	39
<i>culture</i>	14	33
<i>foreign policy</i>	15	47
<i>environment</i>	17	48
<i>mobility and public works</i>	19	64
<i>employment</i>	21	71
<i>welfare and public health</i>	23	86
<i>housing</i>	25	89
<b>TOTAL</b>	157	520

Table 21: A quantitative assessment of indicator orientation by policy sector

The classes of policy indicators vary for different policy fields. Figure 30 represents the total number of indicators for three classes, i.e. input, output and intake, and effect. We combine output and intake categories since intake is often used as a close proxy for output. For example, the intake indicator number of pupils is used to approximately assess the output of a school. Mobility and public works and culture are more input oriented. In these policy sectors, the resources are the most important issue. In the cultural sector, many doubt whether it is possible at all to measure outputs or effects of cultural activities. Input (i.e. subsidies) remains as the most relevant indicator class. Likewise, mobility and public works have a high input orientation. The most interesting question here seems to be how much money has been spent (mostly in the MP's constituency) rather than what the outputs or the effects of the investments are. Education, foreign policy and mainly welfare/public health and housing have a strong focus on output. Housing in particular is characterized by an almost unique focus on output. The main issue in this policy community is the availability of social housing. The effects of the provision of social housing are almost entirely assessed by means of indicators. There is a relatively

greater number of effect indicators in the policy sector 'environment'. The policy sectors 'employment' and 'internal affairs' have a more balanced focus.

## indicator classes by policy field

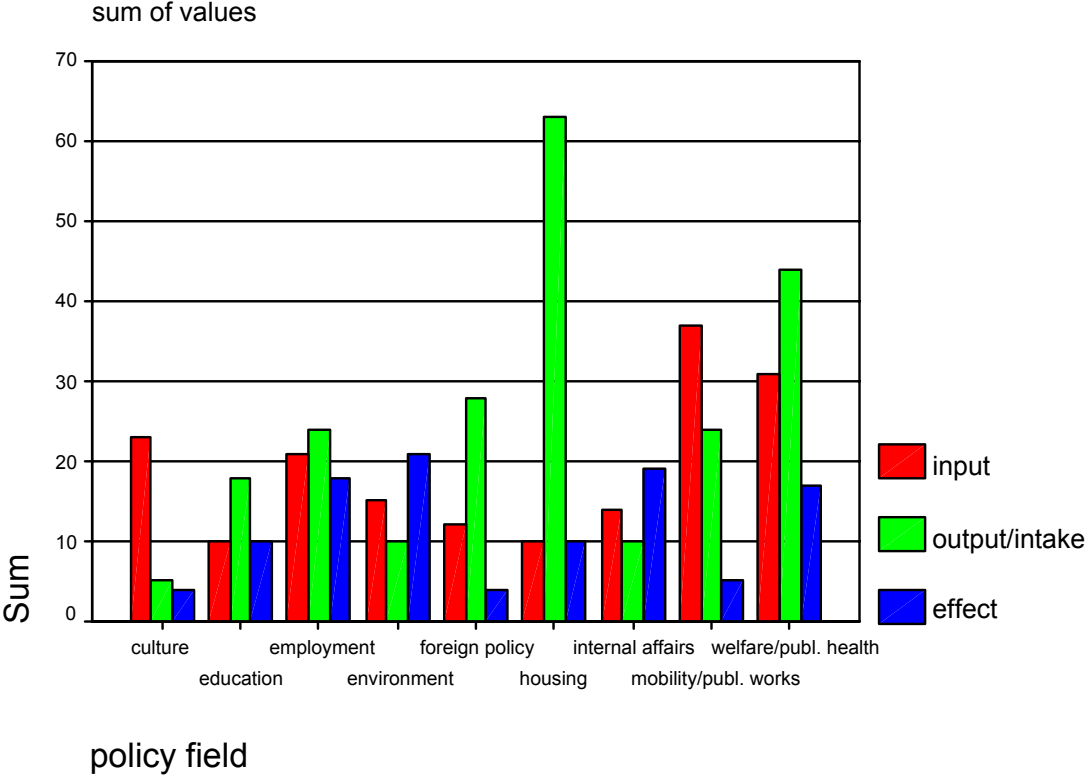


Figure 30: Indicator classes by policy field

The study of the quality of the supply, assessed by the occurrence of direct links in the answer, generates different results compared to the quantitative analyses. For each observation, the ratio between the number of direct links and the total number of demanded indicators has been calculated. This is a quality measure for the answers. A ratio (near to) 1 implies that all the demanded indicators have been provided in a precise way. A ratio near to 0 signifies the opposite. As Figure 31 shows, the policy sectors that have a high score on quantity, do not necessarily have the highest quality. For instance, welfare and public health uses many indicators but does not succeed in providing exactly the demanded indicators. On the contrary, foreign policy and internal affairs use fewer indicators, but when they do so, they are able to provide exactly the demanded indicators. A policy sector that uses indicators in a substantive way does not necessarily imply a better qualitative match. Hypothetically, when a culture favorable to working with indicators is developed, the demand may grow faster than the supply.

## Quality of supply by policy field

Mean (direct link/total demand for indicators)

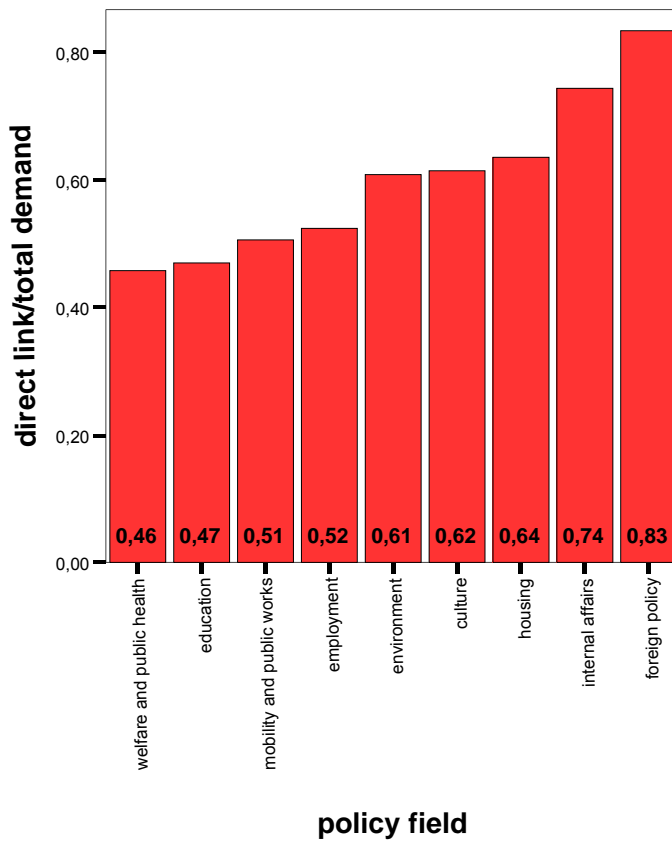


Figure 31: Quality of supply by policy field

By deducting demand from supply, we can compare supply and demand (Figure 32). A negative figure means that demand is higher than supply, i.e. undersupply. A positive figure signifies oversupply. In general, supply is lower than demand. However, when we assess the situation for the different policy sectors and the different classes of indicators (input, output and intake, effect) the picture is more diverse. First, there is considerable oversupply in the policy sector 'employment'. In particular, output indicators are prevalent. The oversupply of output/intake indicators coincides with an undersupply of input and effect indicators. Earlier, we revealed that employment could provide the demanded indicators in only 52% the cases. Nonetheless, the total number of indicators is high. Apparently, output/intake indicators serve as a substitute for lacking input and effect indicators. Secondly, the policy sector *foreign policy* has an oversupply of output and effect indicators. In this case, there is no undersupply. The explanation may be that foreign policy is a relatively small policy sector in Flanders. The impact of one single party is therefore higher. Traditionally, foreign policy is not a precursor of measurement and indicators. It is typical example of a policy sector that is hard to measure. This tradition seemingly reflects in the relative low demand of MP's for indicators. Even so, the Flemish

department of foreign policy made efforts to collect data on output and effect. This initiative of one central organization is presumably reflected in the supply of indicators.

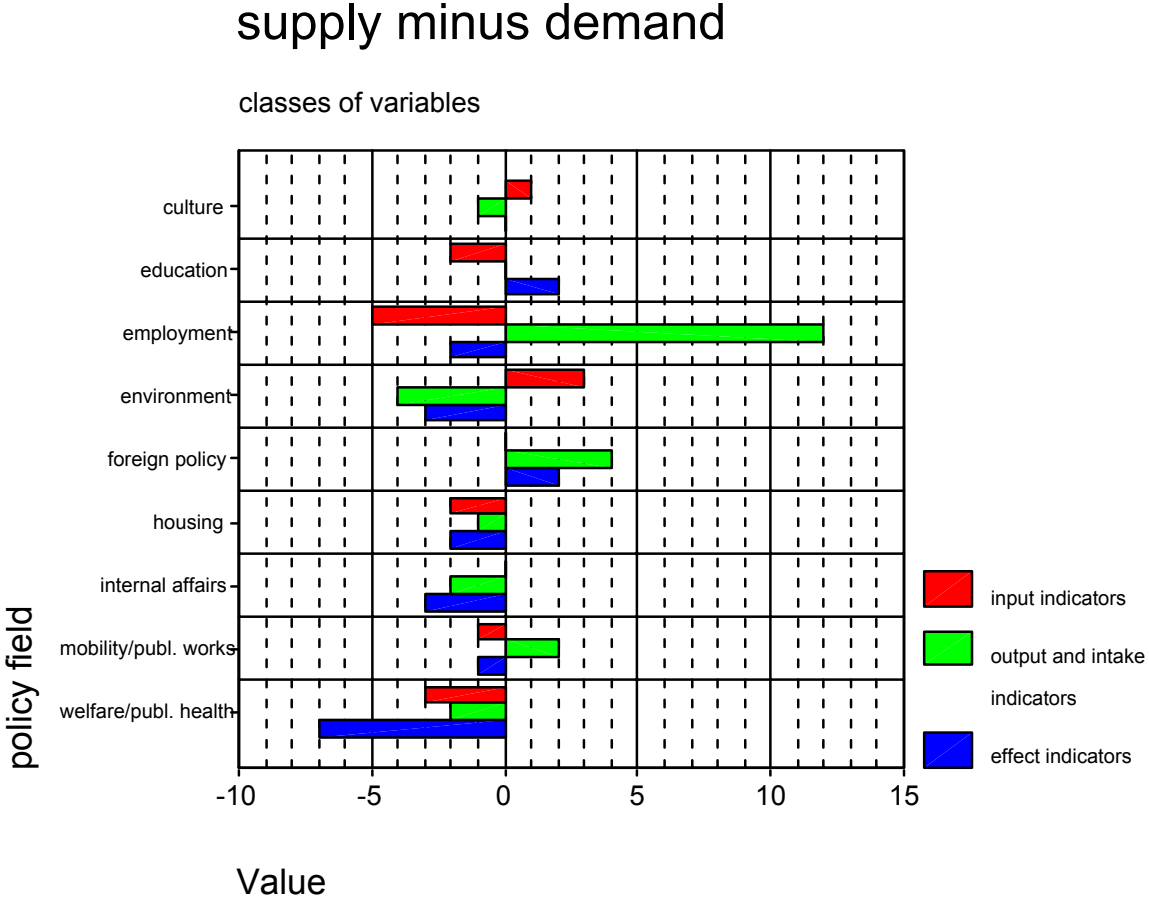


Figure 32: Supply minus demands

Most policy sectors however, show an undersupply of indicators. For instance, the policy sector environment has an undersupply of output and effect indicators. Apparently, an oversupply of input indicators substitutes this undersupply. The most important undersupply can be found in the policy sector 'welfare and public health'. In particular, effect indicators are lacking in this policy sector.

### 7.6. Discussion

The hypothesized differences between policy sectors in the supply and demand of indicators do withstand empirically. The next step is to uncover characteristics of policy sectors that may explain the differences. Following an inductive logic, some explanations may be put forward (Hay 2002). First, the structure of a policy sector may be important. Dominant organizations in a policy community develop an indicator system and have a considerable impact upon the supply of indicators. An example is the

considerable oversupply of output indicators in the policy sector 'employment'. The central employment office developed an extensive management information system to steer and benchmark the local agencies. This management information filters through in the politico-administrative relations. A policy sector without dominant organizations and with a high level of fragmentation may have more consolidation problems and therefore have undersupply. For example, 'welfare and public health' consists of hospitals, facilities for the handicapped, the elderly, the problem youth, the mentally ill, etc. Moreover, both public and private actors manage these facilities. The result is a significant undersupply, especially for effect indicators. However, highly fragmented sectors such as education (schools) and housing (housing corporations) do not have the same undersupply. Thus, other characteristics may come into play such as the prevalence of specific types of issues.

The issue typologies literature argues that clusters of distinct issues have a particular impact on the political relations between individuals, groups and the state. In relation to the indicator orientation of policy communities, this would imply that the dominance of certain issues in a policy sector shapes the relations between actors, i.e. the degree of indicator orientation. An example of a typology is Hogwood's (1978) distinction between principle issues (right or wrong), lumpy issues (involving goods that cannot be divided up), cuts and redistributions (who gets what, who gets more and who gets less?), and increases (who wins more?). We expect less indicators in principle issues, e.g. moral, religious and constitutional matters. For instance, internal affairs mostly deals with the organization of the lower tiers of government (i.e. provinces and municipalities) and the legislation on the use of Dutch as an official language. While these are mainly constitutional matters, this might explain the relatively low number of parliamentary questions with indicators. For lumpy issues, an input orientation may be expected. The sole option open to government is to change the location. This could explain the high number of input indicators for mobility and public works, since this policy sector is mainly concerned with investment in infrastructure. Cuts, redistribution and increase issues have considerably more options. In particular, sectors with a dominance of increase issues may measure more. The increments have to be distributed between parties. This often entails a search for distribution criteria. Therefore, we expect all types of indicators in the sectors with a predominance of these issues. The indicators can play a role in the negotiations. Examples are housing, social welfare and public health.

Thirdly, the measurability of the dominant output that the public sector provides in a policy sector may explain differences in the supply of information. Hackman and Oldman (1980) distinguish between tasks with high routine and low ambiguity, tasks with average routine and ambiguity and tasks with a high routine and ambiguity. The measurement base is respectively time, group and progress. Measurement becomes increasingly difficult when routine lowers and ambiguity rises. In sectors with a high ambiguity and a low routine, measurement of output is more difficult. Typical examples are foreign policy and culture. In the case of the policy sector culture, output and effect indicators are nearly non-existent. To a certain extent, a relatively high level of input indicators substitutes this lack of output and effect indicators. In sectors with a low ambiguity and high routine allow for more output

measurement. Housing is a typical example of a sector with tangible outputs, i.e. the provision of social housing.

Finally, other factors related to the culture and established practices in a policy field may explain differences between policy sectors. These explanations refer to Benson's deep structure of a policy field. In some sectors, indicators are believed to be essential for good policies. Indicators are taken for granted. For example, the unemployment rate is highly institutionalized. It is the single most important statistic and outcome indicator for employment policies. The same applies for GDP growth and several other economic statistics (McRae 1985). Other policy sectors have traditionally less indicators. Notably the cultural sector in Flanders is more reluctant as far as the measurement of output and effect is concerned (Bouckaert and Van Dooren 2000). For outsiders, attendance seems a plain indicator for cultural participation. However, this indicator is contested - perhaps rightly so.

This culturally determined 'way of doing things' might be strengthened through institutional processes such as mimetic and normative isomorphism (DiMaggio and Powell 1983). In conditions of uncertainty, organizations tend to copy the solutions of the organizations in the policy sector that are considered the most successful. When the dominant organizations rely on indicators in their policy and management processes, then the latter practice is more likely to spread. This is termed mimetic isomorphism. Next, normative isomorphism may play a role. Policy sectors usually employ predominantly people with a similar educational background, e.g. arts and humanities in the cultural sector or engineers in the policy sector 'public works and mobility'. It may be that a positive attitude towards indicators is the result of the normative framework that defines what good policy and management are. This normative framework is largely formed through education.

## **7.7. Conclusion**

Let us return to the research questions. First, to what extent is there a supply and demand of indicators? Civil servants often complain about the lack of political interest in the policy indicators that the administration makes available. As a result, they often question the relevance of providing the indicators. The documentary analysis of 270 parliamentary questions presented above however does not confirm this scenario. MP's often ask for indicators on policy issues. Indicators do play an important role in the political arena. Indicators are included in 52% of the MPs' questions to the executive and in 48% of the executive's answers.

Second, what is the quality of the supply? Although there is quantity, the quality of the answers is more uncertain. Only 44% of the answers (with indicators) provide exactly those indicators that have been asked for. The four main problems that were explicitly mentioned are the availability of the data, the lack of timeliness, coordination problems, and consolidation problems. Ideally, politicians should be involved in the measurement process earlier in order to formulate the demand before the

measurement itself takes place. If the political demand is better known beforehand, the data can be made available on time. Consolidation and coordination issues can be anticipated.

Thirdly, are the differences between policy sectors significant and if so, which policy sectors measure more? There are significant differences between policy sectors. In quantitative terms, the policy sectors 'mobility and public works', 'employment', 'welfare and public health' and 'housing' have more indicators and the policy sectors 'internal affairs', 'education' and 'culture' have fewer indicators. In addition, these policy sectors differ in focus of the measurement. For instance, 'Mobility and public works' is input oriented, 'housing' is output oriented and 'environment' is more effect-oriented. Finally, in most policy sectors demand is higher than supply. However, foreign policy and employment show a significant oversupply.

The reason for these differences may be diverse. Structural features such as the extent of decentralization and fragmentation in a policy sector may explain the supply and demand function. In addition, the measurability of the main policy issues and outputs may be a second reason. Finally, different cultural factors and practices in distinct policy sectors may relate to a dissimilar indicator orientation. These different explanations may interact and reinforce each other over time.

Finally, the relevance of the policy sector as a level of analysis stands out. The policy sector is an important unit of analysis, situated between the individual and the organization on the one hand, and the government-wide level on the other.





## 8. What are the system requirements for different uses of performance information?

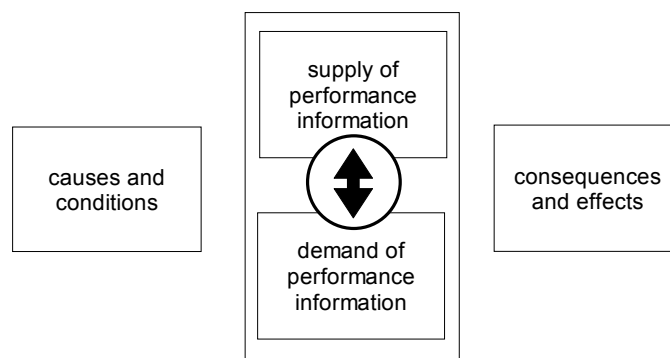
### SUMMARY OUTLINE

**Issue:** one of the most important research questions in performance measurement is “*how to design measurement systems that are ‘fit for use’?*”. However, often the question should be supplemented with “*fit for which use?*”. Behn’s article “*Why Measure Performance? Different Purposes Require Different Measures*” (2003), relates uses to the choice of the indicators. Here, we will relate uses to other steps in the measurement process such as for instance target setting, analysis techniques and reporting

**Research questions:** Two research questions will be addressed. First, what are the uses of performance information? Use is of the independent variable in this research setting. Therefore, we need a classification of uses that goes beyond the techniques and instruments listed on page 36. Secondly, how does use relate to design? How to design a measurement system that is fit for purpose?

**Methodology:** unstructured interviews (N=27)

### Graphical representation



## 8.1. Introduction

The benefit of measurement lies in its use. Few will contest this assertion. There are several uses of performance information. A review of 10 texts on performance measurement (p. 36) yields a list of 44 uses of performance information (Mayston 1985, Osborne & Gaebler 1993, GASB 1997, Hatry 1999, Wang & Berman 2000, Broom et al 2000, GAO 2000, OECD 2003, GPP 2003, Poister 2003). Seldom however, specific uses are retraced to the measurement process that has to provide the performance information. One exception is Behn's "*Why Measure Performance? Different Purposes Require Different Measures*" (2003). He relates purposes of performance information to the choice of indicators.

Yet, choosing the indicators is only one choice in a measurement design. Other choices have to be made. The following is an extended list of issues in designing a measurement system. 1) How will the measurement subject be defined?, 2) What type of indicators will be selected? 3) How will the data be collected? 4) How will be assessed whether the result is good or not?, 5) What analysis techniques will be applied and, 6) What will be the reporting format? The answers to these questions should be determined by the intended use of the performance information. This section seeks to explore this link between use of performance information and measurement system design.

First, we will provide a categorization of the uses of performance information. We need this categories to discuss the implications of the different uses on the main choices in measurement system design in the next section. Secondly, we will discuss the implications of the different uses for the measurement process. An important issue for the conclusions will be the compatibility of different uses of performance measurement information. If the different purpose require different measurement designs, can performance information be used for different purposes at the same time?

## 8.2. Methodology

This chapter relies heavily on unstructured interviews with civil servants in the Ministry of the Flemish Community. We made a purposive selection of 'measurement rich' cases. We assumed that we would learn the most from those cases that have experiences with measurement. Our cases were embedded in the Ministry of the Flemish Community. We used the formal hierarchy for the selection of the cases. We addressed the upper level of the Ministry of the Flemish Community, i.e. the 7 departments. We asked the Secretaries General of the Departments and the Directors General of the Administrations to identify the measurement practices. We interviewed 27 civil servants in 17 sections of the Ministry of the Flemish Community. The sections were both in line and support functions. We invited the respondents to discuss on issues about the history of measurement in their organization, the implementation, the indicators, the analysis of the indicators, and the use of the information.

### 8.3. Uses of performance information

Performance measurement is an organizational process that yields performance information. This information may be used in several ways. Behn (2003) identifies eight purposes of measuring performance (Table 22). He stresses that these are managerial purposes. Politicians and citizens may also use performance information, but these are not subject of his study. In this study, we also focus on managerial use. Administrative supply and political demand has been dealt with in the previous chapter.

<i>Eight purposes that public managers have for measuring performance</i>	
<b>Purpose</b>	<b>The question that the measure can help answer</b>
Evaluate	How well is my organization performing?
Control	How can I ensure that my subordinates are doing the right things?
Budget	On what programs, people, or projects should my organization spend money?
Motivate	How can I motivate line staff, middle managers, nonprofit and for-profit collaborators, stakeholders and citizens to do the things necessary to improve performance?
Promote	How can I convince political superiors, legislators, stakeholders, journalists, and citizens that my organization is doing a good job?
Celebrate	What accomplishments are worthy of the important organizational ritual of celebrating success?
Learn	Why is what working or not working?
Improve	What exactly should we do differently to improve performance?

Table 22: Eight purposes that public managers have for measuring performance

The first purpose is to evaluate. The question is how well the organization is performing. In Osborne and Gaebler (1992) phrase this as *‘if you don’t measure results, you can’t tell success from failure’*. Behn (2003) states that the evaluation purpose is often assumed. People rarely affirm that their dominant rationale is to evaluate. It is between the lines of many performance audits, budget documents, articles, speeches and books (p588). Behn (p.589) feels that evaluation probably is the most established purpose of performance measurement. Whether scholars, analysts or managers like it or not, almost any performance measure can and will be used to evaluate an organizations performance. This is most likely inevitable and even desirable. Innes (1975) statement that the only way a statistician can stay out of politics is to collect irrelevant data largely applies to those who are measuring performance. A crucial question however is who will be doing the evaluating? Will the organization do it itself, or will external actors do it for them?

Secondly, the purpose of performance measurement may be to control. Behn’s main point is that notwithstanding the fact that managers are into empowerment, control is still a main purpose for performance measurement. Although nobody believes anymore in the Taylorist approach of finding

and enforcing one best way of work organization, there is still a control dimension in many measurement systems. This is for instance apparent in the NPM reforms where performance contracts purpose to combine managerial freedom with control over the organization (Kettl 1997). Berman (2002) asserts that even managers whose organizations are not good at performance measurement like it because it gives them control over programs and executives.

The third purpose is to budget. On what programs, people or projects should government spend the public's money? Performance budgeting has become an umbrella for different concepts and practices with regard to the budget format and the budget process (Sterck and Scheers 2006). Firstly, the format may be changed in order to include more performance information<sup>60</sup>. The most far-reaching change is to substitute line item expenditures for performance targets. Subsequently, the line items expenditures are left to managers. Less ambitious is the inclusion of performance information in accompanying documents to the budget. Secondly, performance budgeting may also refer to the process of budgeting. One way of integrating performance information in the budget process is through budget requests. Budget request that attempt to predict outputs may have a greater chance of surviving the deliberation process. Thirdly, performance information may be used for the allocation of resources within line management. Micro allocation decisions may be based on performance data. This decision may comprehend monetary inputs, but also other production factors such as personnel or infrastructure. This would imply that from this micro-perspective, the purpose 'to budget' can be broadened to the purpose 'to allocate resources'.

Fourthly, performance information may be used to motivate people to work towards a target. Osborne and Gaebler (1992) state that *'if you can't see success, you can't reward it'*. Behn points to Compstat within the New York Police Department as an example of the motivational power of performance targets. Crime rates were significantly reduced in New York after introducing a performance measurement system. It should be noted that these measures were accompanied by major reform of the department (Silverman 2001). The motivational power comes from a sense of mission that is shared by the people within the organization. Of course, a mission is not necessarily expressed in a quantitative way. On the contrary, mission statements usually excel in vagueness. Presumably, this elusiveness reduces the capacity of a mission to mobilize people. Quantitative performance targets on

---

<sup>60</sup> There is controversy whether the changes in format will affect budgetary discourse and practice and will really lead to a results orientation (Sterck and Scheers 2006). Mosher (1968) stated that 'the way in which information is classified importantly affects the kinds of treatments and the kinds of decisions that can be made at various levels because the classification framework conditions our subsequent perspectives, understandings and decisions made within the framework'. Other authors point to mainly political variables that they see as more important (e.g. Grizzle 1986).

the contrary have at least a concrete appearance. They may be symbols of this mission - a translation of big words in concrete measurable terms.

The fifth purpose is to promote the organization. A revealing equivocation has developed with the word performance. Performance means achieving results but also showing results. In the same way, public managers do not only have to perform well, but they also have to show the world that they are performing. Performance measures may be purposeful for this aim. Public manager have to convince people that his or her organization is well performing. Osborne and Gaebler (1992) refer to this use : *'if you can demonstrate results, you can win public support'*. Real performance can be reflected by giving a qualitative account of the organization's performance. However, unless there are authoritative sources to underpin the claim, the propensity of a high impact is low. Why would journalists, politicians, interest groups and the like believe the manager? Quantitative measurement by means of performance indicators has the potential of supplying this evidence. It allows to refer to commonly understood phenomena and to agree on their levels (Innes 1990). Measurement does not only used for the definition of problems, but also for the definition of success<sup>61</sup>. Definitely, one of the strategies to promote the organization may be through defining success by measurement<sup>62</sup>.

The sixth purpose is to celebrate. Celebrating accomplishments may lead to a better morale in an organization. It is particularly a complement to the motivational purpose. When indicators are used to make a mission more concrete, shared and understandable, steps towards this mission should also be celebrated. Two differentiations need to be made. First, obviously, celebrating improved performance is only possible when performance is improving. When performance deteriorates, the legitimacy of the measures may decline too. In the end, even the mission and the morale within the organization may erode<sup>63</sup>. Secondly, organizations may celebrate other occasions than a good measured performance

---

<sup>61</sup> The definition of problems and successes through measurement is often considered problematic. Reality is as a rule more complex and many factors are not accounted for. Best for instance (2004) describes several ways in which numbers confuse public issues. Debates are distorted by missing numbers (e.g. selective counting), confusing numbers (e.g. bad graphs), scary numbers (bad news drives out good news), authoritative numbers (high confidence numbers), magical numbers (substitutes for decisions) and contentious numbers (junk science).

<sup>62</sup> Performance measurement is only one strategy in defining success and failure. The opposite of measurement is relying on the power of examples. However, in general, media have a higher tendency to use examples as a way of defining problems, while agencies tend to use quantitative data and compliance with regulation. The distribution of the instruments actors have in defining problems and successes will significantly differ according to the societal context. In Anglo-Saxon countries, league tables are highly publicized. This is not as much the case in Scandinavian countries.

<sup>63</sup> Chris Game (2005) addressed this issue at the EGPA meeting in Berne. The Audit Commission will raise the bar for achieving the status of excellent local community. This will involve a downgrading of a significant number of communities and will probably lead to demoralization.

too. For instance, an increase in budgets or personnel is usually interpreted as a success for the organization. Positive media coverage of the organization's activities is also a reason to celebrate.

Seventhly, organization may measure performance to learn *why* what is working or not. The purpose 'to evaluate' is to answer the *what* question. To learn implies answering the *why* question. Osborne and Gaebler (1992) state that *'if you can't see success, you can't learn from it'*. Yet, data does not speak for itself. If we think they do, this is because 'we each have buried in our brain some unconscious mechanism that has already made an implicit conversion of the abstract data into meaningful information (Behn 2003: p. 592). This conversion, sometimes implicit and sometimes explicit, requires causal models that enlighten the black box. What is happening inside the organizational black box that converts inputs to outputs and what is happening inside the societal black box that converts outputs to outcomes?

Finally, the purpose to improve refers to changes in the operations of the organization. Many authors stress the importance of improvement based on measurement (e.g. Hatry 1999; Poister 2003). Osborne and Gaebler (1992) noticed that *'if you can't recognize failure, you can't correct it'*. Improve is the ultimate aim of measurement. The seven other purposes are means to this end (Behn 2003: p.592). The connection between measurement and improvement however is often not so clear. Who has to do what to convert measurement into improvement? March and Olsen (1976) give some insight into the process of organizational learning<sup>64</sup>. Bounded rational individuals in organizations, with proper beliefs and preferences, receive information from the measurement system. Depending on their capacity to process this information together with other information sources, they translate the interpreted information into individual action. Next, this individual action needs to trigger organizational action. Somebody has to act consciously to alter operations or policies for the better.

The distinction that Behn proposes is a good heuristic device. The questions are sufficiently concrete for empirical research. At this point however, it is time to make explicit an important assumption in this study. We assume that the potential impact on the degrees of freedom of the organization and within the organization is the main distinguishing factor between uses. Degrees of freedom can refer to, for instance, the autonomy to make management decisions, the involvement in strategic and policy decisions and the discretion in implementing policies. Finances and personnel are important enablers for securing these degrees of freedom. Externally available performance information is has potentially a higher impact than information that is kept within the organization. In the same vein, the use of

---

<sup>64</sup> The cycle of organisational learning consists of four steps; a) individual beliefs and preferences determine b) individual action. Individual action determines c) organizational action. Organizational action triggers d) environmental response (feedback) which again influences a) individual beliefs and consequences. Every step may disturb organizational learning results.

performance information may potentially impact the degrees of freedom of individuals and teams within the organization. We think of impact as potential impact. It not necessarily materializes. Conceptually, potential impact is a quantity ranging from high to low. We will review the eight purposes formulated by Behn according to the potential impact they may have on the degrees of freedom of and within the organization.

This is an institutionalist perspective (Scott 1995). Organizations are constrained by institutions. These institutions however are not a given. Institutions themselves arise, persist and diffuse due to actions by organizations. When performance information is used externally, it will either enable or constrain the courses of action of an organization. First, this process may be instrumental, according to a consequential logic (March and Olsen 2004). This is the case when good performance is rewarded and bad performance is sanctioned<sup>65</sup>. In this situation, performance indicators define performance. By doing this, the conceptual distinction between conventional legislative regulation and regulation through performance indicators fades (Laegreid, Roness, Rubeckson 2005). Second, the process may follow the logic of appropriateness (March and Olsen 2004). Actors seek to fulfill the obligations of being part of a social collectivity. They interpret what they think is appropriate in a particular institutional context. By means of performance indicators, the environment of an organization defines and redefines what constitutes appropriate action for the organization.

In a symposium issue of public productivity and management, four authors independently make the distinction between performance measurement for improvement and performance measurement for accountability (Ammons 2002; Berman 2002, Hatry 2002; Halachmi 2002). The Working Party on Performance Monitoring in the Public Services of the Royal Statistical Society (2005) adds a third function - the managerial role. The eight purposes characterized by Behn could be seen as subdivisions of the classification of the Royal Statistical Society. Three functions are identified: research, internal management and accountability.

1. First, may performance information be collected in order to find out what works? This is the research function. The main rationale is learning. The rise of evidence based policy initiatives may be seen in this light (Davies, Nutley, and Smith 2003; Boaz and Nutley 2003). The underlying assumption on human and organizational behavior is generally a positive one. McGregor would typify it as Theory Y (1960). People are intrinsically motivated to perform well and to seek responsibility. This also motivates them learn how to do the job. How can policy or management be improved? Policy planning

---

<sup>65</sup> Rewards and sanctions may be of many kinds. In the Niskanen tradition, the valued rewards and sanctions are increases and cuts in budget and personnel (1971). Dunleavy (1991) reviewed this theory and proposed 'bureau shaping' opportunities as the reward and sanction system. His central argument is that bureaucrats strive towards prestige rather than mere budget maximization.

and evaluation, strategic planning and evaluation and business process re-engineering are examples of policy and management tools with a primarily research orientation. Performance information can be used for process evaluation and outcome evaluation which envisages, respectively, service improvement as policy improvement (Scheirer 1994).

2. Secondly, the internal managerial function of performance information is about identifying and sanctioning well performing or underperforming institutions or public servants (to control, to motivate) and about allocating resources (to budget). The purpose is not to learn from mistakes, but to agree upon targets. The underlying assumption about human and organizational behavior is generally a negative one - Theory X in McGregor's words (1960). Theory X assumes that the prime source of most employee motivation is monetary, with security as a strong second. People will be shirking and gaming in order to serve their own self interest. Performance management tools need to counter this behavior<sup>66</sup>. The use of performance information for management tools such as internal performance contracts, performance based pay, performance budgeting and performance mandates fall in this category (Bovaird and Loffler 2003; Kamensky and Morales 2005).

3. The third purpose is accountability. The main proposition is that the public sector should be accountable to the citizens/taxpayers and politicians<sup>67</sup>. Therefore, performance of public bureaucracies should be disclosed to the public. In particular in the Anglo Saxon countries, the accountability culture is strong (Dubnik 1998; Ammons 2002). The underlying assumption bears resemblance to the managerial role, but in this case the pressure is coming from outside of the organization. Two categories of external pressure can be identified. First, pressure is assumed by showing results to the general public. In case of a monopoly, the potential criticism of the public is expected to exert enough pressure on the organization. In case of (quasi-) markets, for instance in public schools and hospitals, market pressures are provoked by publicizing rankings (the so called league tables). Typically, league tables, citizen charters and upgraded annual reporting are examples of this accountability role (Bowerman 1995; Gormley and Weimer 1999). Secondly, pressure is instituted by the political system. Performance contracts with agencies are a good example. These contracts give autonomy to agencies provided the agency commits itself to output or outcome targets within the budgetary framework.

---

<sup>66</sup> The more mild expressions of this approach still see people as pursuing self-interest, but they define it broadly as what the one finds important. Altruism then is also considered to be one's self interest - for instance because it gives peoples 'a warm glow' (Andreoni 1990).

<sup>67</sup> Aucoin and Heintzman (2000) identify three purposes of accountability. The first is to control for the abuse and misuse of public authority. The second is to provide assurance in respect to the use of public resources and adherence to the law and public service values. The third is to encourage and promote learning in pursuit of continuous improvement in governance and public management.



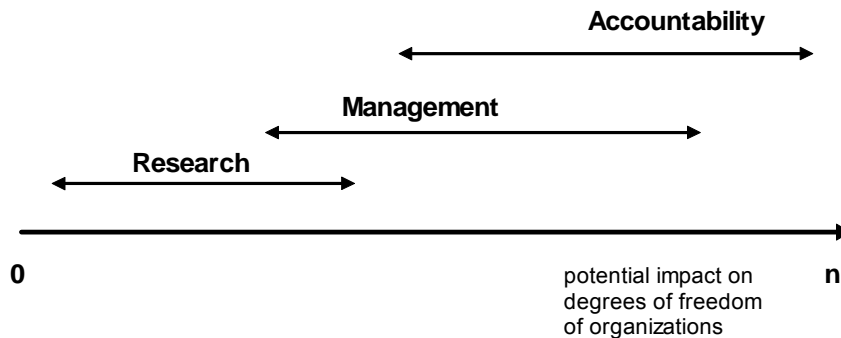


Figure 33: Potential impact of the use of performance information on degrees of freedom of organizations.

Figure 33 gives an indication of the different potential of the use of performance information to impact the degrees of freedom of and within organizations.

1. In general, the research function has the least direct impact on the degrees of freedom. On the contrary, the learning organization is one in which everyone is engaged in a problem solving (Senge 1990). People and leaders have a shared commitment to learning and believe they can change their environment. There is open and extensive communication. Also, there needs to be some slack time to give occasion for learning. Surely, this is an idealistic view on organization. Undoubtedly, public managers will take the potential negative consequences of evaluation and learning for the organization into account. When the evaluation of a program proves it is flawed, it may have an impact on organizations. However, this impact varies according to the context in which the organization operates. The double-sided arrow in Figure 33 represents the fact that even within the research function, there is a range of different positions.

The purposes 'to evaluate', 'to learn' and 'to improve' fall mainly into the research category. To improve is the ultimate objective of all uses of performance information. However, since the main question is '*what exactly should we do differently to improve performance?*' (Behn 2003), it is included within the research function. It is the logical next step after evaluating and learning. The potential for improvement results from the measurement effort. This is different for the other uses (to budget, to control, to motivate, to promote, and to celebrate) where there is an intermediary factor between measurement and improvement, i.e. the behavioral responses to the use of performance information.

2. The management function potentially has a higher impact on the degrees of freedom within the organization. Performance information plays an important role in management control (Rejc 2005). Management control is the process of guiding organizations into viable patterns of activity in a changing environment (Berry, Broadbent and Otley 1995). Managers have to assure that resources are obtained and used effectively and efficiently to achieve the organizational goals (Anthony 1965). This implies that control is about influencing "the behavior of organizational participants so that some

overall organizational goals are achieved” (Berry, Broadbent and Otley 1995: p4). Performance management systems are widely used for internal control purposes. Many examples can be found in the management accounting literature where control is seen as a primary purpose of performance information (e.g. Adnum 1993; Cavalluzzo and Ittner 2004). Whereas the research function presupposes an intrinsic motivation of people to contribute to the organization’s goals, the internal management function starts from an extrinsic motivation that has to align with the goals of the orientation by means of a good control system.

The purposes that relate to internal management are firstly ‘to control’ and ‘to motivate’. The question is how to ensure subordinates are doing the right things and to motivate line staff, middle managers, nonprofit and for-profit collaborators, stakeholders and citizens to do the things necessary to improve performance<sup>68</sup>. Performance pay is an example of a management tool in this category. Secondly, the purpose ‘to budget’ is also an internal management purpose – at least at the micro level. The question is on what programs, people, or projects the organization should spend money?

3. The accountability function potentially has the biggest impact on the organization’s degrees of freedom. In the most extreme form, performance information becomes a matter of life or death, or at least of thriving or coping. This is for instance the case when bad scores in a league table leads to changes in the quantity and characteristics of the intake of a school (Bradley, S., Crouchley, R, Millington, J., and Taylor, J. 2000; Wiggins and Tymms; 2002 Wilson 2004). Equally, when an organization does not attain the targets in a performance contract, sanctions may be severe - in particular for the management of the organization. The board of the Dutch railways for instance was dismissed in 2002 because the organization did not succeed in making 80% of the trains run on time. The actual number was 79.9%. Performance targets served as the staff to beat the dog (Vrij Nederland 19 januari 2002). The president of the Dutch railways claimed that 0.1% of target was not bad. Politicians claimed that 80% was already a low minimum and that passengers were dissatisfied with the services (Het Parool 3 January 2002). Clearly, performance targets are not the only one factor in the evaluation of the agency. By no means, a bad score on the indicators leads univocally to sanctions<sup>69</sup>. The main ‘threat’ arises from the possibility to use indicators of the performance contract against the organization.

The purposes ‘to promote’ and ‘to celebrate’ fall into this category. The former is about convincing

---

<sup>68</sup> Non-profit, for profit, stakeholders and citizens regularly are co-producers of public services (Van Dooren, Bouckaert and Thys 2004). This is the role they play in the internal management.

<sup>69</sup> The sanctioning of agencies that do not achieve targets is a sticky one. Should underperformers be punished, which would make it even more difficult to live up to the performance expectations? Or, should underperformers be

political superiors, legislators, stakeholders, journalists, and citizens that the organization is doing a good job? The purpose to celebrate refers to the accomplishments that are worthy of the important organizational ritual of celebrating success? These are positive formulations from the manager's perspective of the accountability function. Often, external actors such as the media or the legislature enforce this accountability.

Behn 2003	Royal Statistical Society (2003)
Evaluate	Research function
Control	Management function
Budget	Management function
Motivate	Management function
Promote	Accountability function
Celebrate	Accountability function
Learn	Research function
Improve	Research function

Figure 34: Transition table from eight management purposes (Behn 2003) of performance information to three functions of performance information (Royal Statistical Society 2003).

**8.4. The measurement process with a differentiation for use**

We now turn to the supply side of performance information. The supply side is conceptualized as a process that yields performance information (Hatry 1999; 2003). We will look at the steps in a typical measurement process and discuss how the parameters for design relate to the three uses.

**8.4.1. Step 1 establishing the implementation mode of measurement**

The first decision in the design of a measurement system is the implementation mode. This involves two choices. First, there is the choice for a bottom-up or a top-down approach to implementation. The top-down method focuses on implementation of centrally controlled variables and deemphasizes other factors. The bottom-up approach has been to start with action at the "street level" and then to generalize the practice throughout the organization (Sabatier 1986; O'Toole, 1989, p.2)<sup>70</sup>. Whereas a top-down approach usually has the advantage of uniformity and less coordination cost, the bottom-up

---

provided with more resources in order to improve, which may do injustice to other organizations that did meet the targets?

<sup>70</sup> The concepts top-down and bottom-up are originally formulated as approaches to implementation research (O'Toole 1989). Some authors look at formal policy mandate to which actors have to adapt. It is a design-driven approach. Other researchers look at the complex street level variation in values, strategies and tactics. Yet, the two approaches are complementary rather than contradictory (Sabatier 1986). They explain parts of reality. Therefore, they may be seen as approaches to implementation as well as to implementation research.

approach has the advantage of a higher ownership and acceptance within the organization. A total bottom-up or top-down approach may be seen as extremes on a continuum. There will always be some top-down and some bottom-up elements in the implementation of a system. The issue is what proportion of each is desirable.

Secondly, the implementation mode may be incremental, consist of a single package or be decremental. Figure 36 represents the implementation strategies. Incrementalist approaches will introduce the measurement system in small packages. New practices are only marginally different from past practice. This is a well-known concept in policy sciences. Incrementalists are continually building out from the current situation, step-by-step and by small degrees (Lindblom 1959). A single package approach will implement the whole measurement system at once. This approach compares to the rational-technical approach in implementation literature (Parsons 1995). With the term decremental approaches, we point to the instance where an organization implements a single package system and gets rid of aspects of the measurement system throughout time. This is a familiar observation with measurement systems. The U.K. audit commission for instance drastically reduced the number of indicators in its measurement system throughout time (Bouckaert Depeuter and Van Dooren 2004). The decomposition of the measurement system takes place in small steps. The decremental approach is indeed than a variety of instrumentalism.

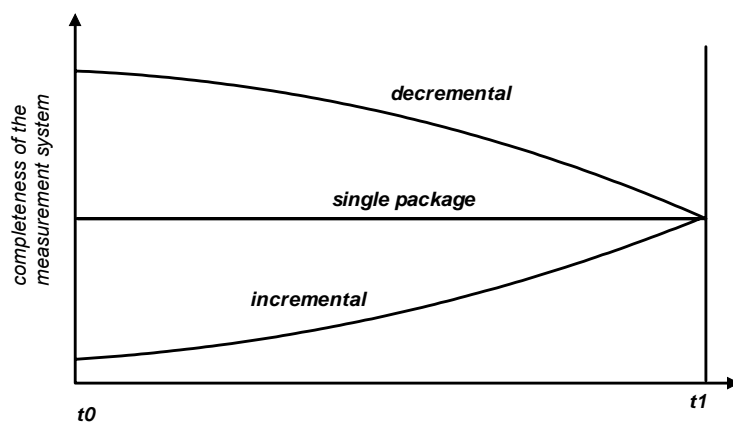


Figure 35: Implementation strategies

There is a trade off between incremental, decremental and single package approaches. The advantage of a single package is that the 'reform stress' on the organization is concentrated in time. The downside is that resistances as monetary costs are also concentrated in time. With an incremental approach, the uncertainty as the cost and resistance are spread throughout time. A decremental approach has potentially decreasing stress levels, since the bits that are not working are abolished. Costs and resistance will also decrease provided that the malfunctioning of the measurement system is the criterion for abolishment and not political or bureau-political tactics. A

decremental approach implies the availability of slack resources. Since it is a trial and error approach, initial investment in building a system will, after some time, prove to be ineffectual.

8.4.2. Differentiation for purposes of performance information

The research function is a low-pressure function. A bottom -up and incremental approach may be the best option. The legitimacy inside the organization of measurement will be high. This is a necessary condition for organizational learning. Reform stress and uncertainty are relatively low. As a consequence, the spread of reform stress throughout time is less of a problem.

The management function may benefit from a mild top-down implementation and a single package. Several management tools -for instance rewarding performance or performance contracts- require stability in the measurement system. Therefore, the whole set of indicators needs to be in place when for instance the contracts are signed. The internal management function is rather a control function than a learning function. Uniformity therefore is important. Top-down approaches result in more uniform practices in the organization.

Finally, the accountability function may benefit from a measurement system that initially is too big. Performance information for external accountability potentially puts high pressure on the organization. Moreover, in the eyes of the organization the external judgment usually is overly simplified. Therefore, it may be beneficial to have some slack data to contextualize these judgments. In advance, it is not always clear which indicators external actors will pick up. After some time, the focus of scrutiny by external actors becomes more predictable. This would allow for cutting down on the measurement system and reducing slack measurement. The measurement system will be implemented top-down. The contents of the measurement system will be defined by the external demands rather than internal. It is the responsibility of apex to take care of the institutional relationships of the organization -with politicians, with the media, with interest groups, etc.

<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
bottom-up	rather top-down	top-down
incremental	single package	decremental

Table 23: Differentiation of implementation mode according to the purpose of performance information.

8.4.3. Step 2: definition of the measurement object

The second phase in the measurement process is the definition of the measurement object. Here, the question what to measure is answered. It is pointless pursuing a measurement system that perfectly mirrors every aspect of the organization, its policies and environment. Thus, a choice has to be made on what to measure and what not to measure. With the definition of the measurement object, the

coverage rate of the measurement system is set. Starting from the input output model (see Figure 3 on p. 28), different cutouts of the organization or the policy sector may be defined.

---

**Which part of the organization will be measured?**

- ✓ Which part of the organization chart? All the services, or only a limited selection?
- ✓ Which input? Which part of the budget, personnel, infrastructure?
- ✓ Which activities? Which policy instruments?
- ✓ Which outputs? Which products of the organizations (goods and services) are being measured?

**Which part of the policy objectives is being measured?**

- ✓ Which intermediate ends? Which target groups? Which geographical circumscriptions?
  - ✓ Which outcomes? Only the intended outcomes, or the side effects ?
  - ✓ Which contextual factors are taken into account?
- 

Table 24: Definition of the measurement object.

Several criteria may be used to motivate the definition of measurement objects. Underlying a first list of criteria is an **efficiency of measurement** perspective. The argumentation primarily refers to the Pareto principle (the 80/20 rule). The 80/20 rule means that in anything a few (20 percent) are vital and many (80 percent) are trivial. With 20% of measurement, 80% of the measurement objective may be attained.

1. *Indications of problems* through symptoms: complaints, waiting lists, and statistical info. With a specific measurement, the most urgent issues in the organization may be mapped.
2. *Financial importance*: in many organizations, a small amount of activities accounts for the bulk of the budget. By measuring these activities, the organization has a good coverage of the budget.
3. *Personnel deployment*: by measuring a limited amount of activities, most of the personnel may be comprised in the measurement system.
4. *Societal visibility*: some activities, which may not have a high financial impact, may still have a high societal visibility, with a lot of interest from the media, politicians and the civil society. By measuring these activities, the organization may be able to respond to most of the issues that those actors raise.

Underlying a second list of criteria for choosing measurement objects is a **developmental perspective**.

5. *Added value* of performance measurement; what information does the organization already have and where does it want to go?
6. *Feasibility* of the measurement effort: in order to overcome resistance and to make people accustomed to measurement, some quick wins may be beneficial. In this case, feasibility is a legitimate criterion. Provided that the organization intentionally follows this strategy. When this strategy is followed subconsciously, this may lead to a tunnel vision on the measurable aspects of the organization.

7. *Distribution of measurement efforts* in the organization: sometimes measurement efforts are distributed throughout an organization. The option to have a little measurement for many rather than a lot measurement for some may for instance be prompted by the desire to introduce a results oriented culture in the whole organization.

Thirdly, sometimes there is no choice. What to measure is **predetermined** externally. This is for instance the case for international reporting obligations (European Union, OECD, etc.). For local and regional governments, central government may impose the definition of the measurement object.

#### 8.4.4. Differentiation for purposes of performance information

Table 25 differentiates the criteria for deciding what to measure for the three main functions of performance information. The research function has a strong focus on learning. Indications of problems therefore are good grounds for measurement. In the previous step, we argued that an incremental implementation is suitable for the research function. Consequently, a developmental perspective may complement this incremental approach. In particular, the added value vis-à-vis the information that is already available and the feasibility of the measurement are good motivations to fulfill the research function.

The measurement object for the internal management function may be defined from a developmental perspective. In particular, the distribution in the organization may be a relevant criterion. Management scorecards prescribe this approach (Kaplan and Norton 2001). Each section of the organization should provide some key indicators in order for the top of the organization to monitor the operations. When the purpose is to motivate or to budget, respectively personnel coverage or financial coverage may be valid motivations for deciding what to measure and what not to measure.

Accountability requirements often are imposed on organizations. In this case, the definition of the measurement object is predetermined. Others decided for the organization what they should measure. If however the organization is taking the initiative to promote itself, social visibility may be the main criterion for selecting the measurement object. An employment agency for instance may want to direct measurement efforts to results in those target groups that are prominent in the policy debate. The intermediate position between complete external determination and voluntary promotion is the negotiation of the measurement object. This is clearly the case in performance contracts between agencies and departments. Also, organization can influence the content of league tables or imposed reporting formats through lobbying.

<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
efficiency of measurement	efficiency of measurement	efficiency of measurement
✓ indications of problems	✓ financial coverage	✓ societal visibility
developmental perspective	✓ personnel coverage	predetermined
✓ added value	developmental perspective	
✓ feasibility	✓ distribution	
✓ distribution		

Table 25: Differentiation of criteria for the definition of the measurement object according to the purpose of performance information.

### 8.4.5. Step 3: selection of the indicators

The third phase is the selection of the indicators. Indicators may be single indicators that refer to inputs, outputs, outcomes or the environment in which the organization operates. Indicators may also be ratio indicators that combine single indicators. Efficiency is input over output. Effectiveness is output over effect. Cost effectiveness is input over effect.

### 8.4.6. Differentiation for purposes of performance information

Table 26 translates Behn's (2003) recommendations to the three functions of the use of performance information. The indicators needed for the research function have to reflect the whole 'production chain of the organization'. Indicators on inputs, outputs, outcomes and factors that may influence the outcome are required. Moreover, they need to sufficiently disaggregated to diagnose deviances from the expected (Behn 2003). A more or less complete picture is optimal. However, measurement will seldom univocally dictate answers. Interpretation of the indicators is required. The room for interpretation allows for qualitative knowledge and experience to fill the gaps in the measurement system.

The indicators for internal management control will mainly have to be a combination of input and output indicators to address the management purposes. Moreover, they should not allow for much interpretation. Outputs usually are better observable than outcomes. Even if outcomes are observable, the causal link with the organization's activities needs to be identifiable.

Optimal indicators for accountability are outcome indicators. Outcomes are the most relevant for society (Hatry 1999). They are the reason of existence for a public organization. However, because of the potential impact on the organization, the indicators should not leave room for too much interpretation. In addition, the causality between the outcome and the organization's activities should be clear. These are qualities which outcome indicators seldom have. Therefore, output indicators are used instead although they may give an incomplete picture with an output bias.



<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
✓ the whole 'production chain of the organization'	✓ mainly output and efficiency ✓ moderate room for interpretation	✓ optimally outcome, with output as a second best
✓ room for interpretation		

Table 26: Differentiation of the choice of the indicators according to the purpose of performance information.

#### 8.4.7. step 4: data collection

After the selection of the indicators, the data can be collected. Self-evidently, this is a crucial step in the process. Organizations may use internal or external data sources. Internal data is produced by the organization itself. External data is purchased or obtained from other organizations. A broad array of data sources may be used. (United Way of America 1996; Wholey, Hatry and Newcomer 1994; Weiss 1998; Hatry 1999; Abravanel 2004). Table 27 assesses the advantages and disadvantages of different data sources.

<i>Data source</i>	<i>Advantages</i>	<i>Disadvantages</i>
Existing registrations	<ul style="list-style-type: none"> <li>✓ continuity</li> <li>✓ low cost</li> <li>✓ in-house, good insight into quality and content</li> <li>✓ readily available</li> </ul>	<ul style="list-style-type: none"> <li>✓ no drop out data</li> <li>✓ less focus on outcome</li> </ul>
Additional registrations	<ul style="list-style-type: none"> <li>✓ continuity</li> <li>✓ in-house, good insight into quality and content</li> </ul>	<ul style="list-style-type: none"> <li>✓ 'hidden' costs</li> <li>✓ medium- to long-term availability<sup>71</sup></li> </ul>
Surveys	<ul style="list-style-type: none"> <li>✓ suitable for outcome information</li> </ul>	<ul style="list-style-type: none"> <li>✓ high cost</li> <li>✓ medium term availability<sup>72</sup></li> <li>✓ response rate issue</li> </ul>
Self-assessments	<ul style="list-style-type: none"> <li>✓ low cost</li> <li>✓ combination of quantitative and qualitative approaches</li> <li>✓ linked to operations</li> </ul>	<ul style="list-style-type: none"> <li>✓ perceptual</li> <li>✓ risk of gaming</li> </ul>
Technical measurement <sup>73</sup>	<ul style="list-style-type: none"> <li>✓ non-obtrusive</li> </ul>	<ul style="list-style-type: none"> <li>✓ limited applicability on human services</li> <li>✓ risk of technocracy</li> </ul>
External observers	<ul style="list-style-type: none"> <li>✓ limited obtrusiveness</li> </ul>	<ul style="list-style-type: none"> <li>✓ high costs for specialized observers</li> </ul>

<sup>71</sup> Additional registrations will yield data more quickly when there is a relatively high number of files and the typical record or file of the organization has a short processing cycle. An employment counselling service for instance will have extra data more quickly compared to a fiscal administration (with typically a one year cycle) or an organization that deals with foreign investment projects.

<sup>72</sup> Polling may yield data in shorter notice, often at the expense of validity and/or reliability.

<sup>73</sup> Applications may be found for instance in the environmental sector (air quality, water quality), in housing (level of humidity as an element of housing quality) and in public health (toxic substances in the population).

<i>Data source</i>	<i>Advantages</i>	<i>Disadvantages</i>
	<ul style="list-style-type: none"> <li>✓ observers are not involved</li> </ul>	<ul style="list-style-type: none"> <li>✓ medium to long term availability</li> </ul>
Other public organizations	<ul style="list-style-type: none"> <li>✓ usually low cost</li> <li>✓ short term availability</li> </ul>	<ul style="list-style-type: none"> <li>✓ confidentiality and privacy issues may interfere data exchange</li> <li>✓ less insight into quality and content (definitions)</li> </ul>
Statistical, international, and research institutions.	<ul style="list-style-type: none"> <li>✓ good quality<sup>74</sup></li> <li>✓ authoritativeness</li> <li>✓ readily available</li> <li>✓ moderate costs</li> <li>✓ continuity</li> </ul>	<ul style="list-style-type: none"> <li>✓ not directly tailored to organization's needs</li> <li>✓ only outcomes</li> </ul>

Table 27: Advantages and disadvantages of different data sources

#### 8.4.8. Differentiation for purposes of performance information

The research function may benefit from the broadest range of data. Existing and additional registrations are useful foundations in order to evaluate, to learn and to improve. Self-assessments can be used to improve internal processes. Technical measurements and surveys can provide insight into the outcome of the organization. Data from other organizations and statistical institutions will primarily be useful to establish causality and the influence of contextual factors on the outcome. External observers are the only data source that may be less useful for the research function. The research function postulates the intrinsic motivation to learn. The benefit of impartiality is less pressing than the need for ownership of the findings by those who will have to learn from it.

The internal management function -to budget, to control and to motivate- will be dependent on existing and additional registrations and record keeping within the organization. Where applicable, technical measurements are useful as a control device thanks to their unobtrusive character. When technical measurement is not feasible, external observers can take over the role of neutral bystander. Reliance on other organization's data, statistical data, and survey data is less useful for internal management since they rather comprehend outcome and contextual information. Self-assessments may also less usable for internal management focuses. They rely heavy on perceptions are relatively easy to manipulate. When the data is used for allocation decisions, there is both an opportunity and the motive.

---

<sup>74</sup> Sometimes it may appear that authoritativeness is used as a substitute for data quality - in particular in the international institutions (Van Dooren, Sterck, and Van De Walle 2005). A review of the European Central Bank data on Public Sector Efficiency, World Bank data on government effectiveness, World Economic Forum data on Public institutions and IMD business school data on government efficiency shows serious weaknesses in all four studies (Van De Walle 2005).

Thirdly, administrative registrations are also useful for the accountability function. If organizations have to show which outputs, target groups and regions they have served, they will primarily look at the own records and files. Additionally, technical measurements, external observers and surveys (for instance about client satisfaction) may be used. The subjective elements in surveys is however a problem for tough accountability relations. Other organization’s data and data from statistical institutions will usually only play a peripheral role in contextualizing success and failure. Finally, self-assessments are not useful for the same reason as for the internal management function.

<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
✓ Existing registrations (+)	✓ Existing registrations (+)	✓ Existing registrations (+)
✓ Additional registrations (+)	✓ Additional registrations (+)	✓ Additional registrations (+)
✓ Surveys (+)	✓ Surveys (-)	✓ Surveys (+)
✓ Self-assessments (+)	✓ Self-assessments (-)	✓ Self-assessments (-)
✓ Technical msmt (+)	✓ Technical msmt (+)	✓ Technical msmt (+)
✓ External observers (-)	✓ External observers (+)	✓ External observers (+)
✓ Other organizations (+)	✓ Other organizations (-)	✓ Other organizations (-)
✓ Statistical, international, and research institutions (+)	✓ Statistical, international, and research institutions (-)	✓ Statistical, international, and research institutions (-)

Table 28: Differentiation of the data sources according to the purpose of performance information. (-; less useful; +; useful)

8.4.9. Step 5 analysis (from performance data to performance information)

Fifthly, the analysis of performance information is about making sense of data. The purpose of analyzing data is to transform data into information. Data is raw material. Information may lead to conclusions. Sometimes raw data already provides some information. An unemployment number for instance already gives some information. However, to fully explore the potential of the information, a good analysis is required.

a. First, is the result good or not? In order to answer this question, the result is confronted with a norm. There needs to be a yardstick. There needs to be a basis to make judgments. (Weiss 1998) This norm may be set in advance. In this case, it is called a target. Where do targets and norms come from? (Table 29: Foundations for targets)<sup>75</sup>. First, targets can be based on time. The norm then usually is to do at least as good as last year. Secondly, norms can be based on comparisons with other organizations; within the sector, outside the sector, or in other countries. Within organizations, sections may be compared. The norm can be the average, the top quartile or the best performing parts. Thirdly, scientists can calculate the norms. The amount of harmful substances that can be tolerated in our food and the environment are examples. Fourthly, norms may have a political

---

<sup>75</sup> In some cases, there is no conscious deliberation about the norm setting. Although it may be argued that the norms are set arbitrarily, scrutiny will probably reveal an implicit frame of reference.

foundation. They are put for the symbolic and appealing character. Absolute norms such as 100% coverage are usually not attainable. However, for symbolic reasons, they are maintained. The message is that government should not rest on his laurels, when for instance the 95% target is attained.

b. The second question is why a particular result is showing. This is about getting insight from data by searching for the causes of observations. What are the variables that explain the result of an indicator? This approach is related not wholly disconnected from breaking out of data. The choice of the breakout categories (for instance ethnicity, gender) often is based on hypotheses about the explanatory variables. However, causality analysis is substantially more far reaching. First, in a causality analysis, the underlying theory is made explicit. Secondly, the relations are also tested. In many cases, the statistical analysis will not be sufficient. In order to get a more profound insight into the causes, qualitative research (e.g. interviews, focus groups, etc.) may be undertaken. Performance information often is scratching the surface. It says what is wrong, but not why. Performance measurement and in depth research are thus complements.

Fundament	Assessment	Example
Time	<ul style="list-style-type: none"> <li>- fit for unique policy initiatives</li> <li>- fit for organizations that have no counterpart</li> <li>- fit for confidential information</li> <li>- contextual variables may cause disturbance</li> <li>- risk of stagnation, no innovative impulses from the outside</li> </ul>	- the number of youth in special care did not rise
Other organizations within the sector	<ul style="list-style-type: none"> <li>- fit for comparing results of policies</li> <li>- learning effects through confrontation with other practices</li> <li>- contextual variables or events have less impact when the organizations are quasi in the same context.</li> </ul>	<ul style="list-style-type: none"> <li>- the stress-index for personnel of different organizations in the public sector</li> <li>- the crime figures of one big city compared to another big city</li> </ul>
Other organizations outside the sector	<ul style="list-style-type: none"> <li>- fit to compare management results</li> <li>- learning effects through confrontation with other practices</li> <li>- comparability is harder to achieve</li> </ul>	- sick leave in the private sector versus the public sector
Foreign organizations	<ul style="list-style-type: none"> <li>- fir for monopolists that have no national counterparts.</li> <li>- learning effects through confrontation with other practices</li> <li>- difficulty of overcoming cultural and structural differences.</li> </ul>	- comparison of the educational achievement through the OECD's 'education at a glance' reports.
Scientific standards	<ul style="list-style-type: none"> <li>- well funded, less debatable</li> <li>- technical, risk for technocracy</li> </ul>	- the vaccination level of the population that should be attained in order to eradicate a disease
Political and ideological norms	<ul style="list-style-type: none"> <li>- embedded in the system, higher acceptance of the whole measurement system</li> <li>- not always realistic, (but not necessary unrealistic)</li> </ul>	The zero norm for traffic casualties.

Table 29: Foundations for targets

c. Thirdly, the organization may ask for who (e.g. for which target groups) and where (e.g. for which areas) a good or a bad result is showing precisely. This will require the breaking out or aggregation of the data to the right level. For some purposes, more detailed information will be needed (for instance for cost accounting). For other purposes, the information may have to be more general at a consolidated level (for instance for reporting to parliament). Different purposes will require differences in detail.

Breaking out and aggregation can be done in two ways: on the measurement objects or on the indicators. The breaking out and the consolidation of information may firstly be oriented towards the measurement object, such as regions or target groups. The indicator ‘traffic casualties’ can be broken out for different regions, or can be consolidated on a national level. Secondly, the breaking out and consolidation may be oriented towards different indicators that say something about one single measurement object. An example is the composition of a quality of life index for a neighborhood. The measurement object remains the same (the neighborhood), but the information on the neighborhood is brought together. Indicators may for instance be the average surface of the houses, the number of crimes per capita, the density of the traffic etcetera. Table 30 gives an example for water quality. The level of aggregation thus may range from a single indicator to an index of indicators on the one hand and from a single unit observation to a multitude of observations on the other hand.

direction indicator subject	indicator oxygen	indicator fish stock	indicator nitrogen..	$\Sigma$ indicators
Measurement subject: River 1	Oxygen in River 1	fish stock River 1	nitrogen in River 1	Water quality in river 1
Measurement subject River X	Oxygen in River X	Fish stock River X	nitrogen in River X	Water quality in river X
Measurement subject River Xn	Oxygen in River Xn	fish stock in River Xn	nitrogen in River Xn	Water quality in river Xn
<b><math>\Sigma</math> measurement subjects</b>	Oxygen in all rivers	Fish stock in all rivers	nitrogen in all rivers	Water quality in all rivers

Table 30: An illustration of the breaking out and aggregation of data.

The methodology for breaking out and consolidating should be made known. Interest groups, politicians, auditors, and other specialists should be capable to assess the way data is consolidated or broken out. Indicators are often suspicious, in particular when the methodology is not stated (Best 2004). On the one hand, positive data can be sought for by breaking out for the right categories. On the other hand, negative data can be presented in a much nicer way by diluting them in a composed measure. Indexes often give different weights to the composing indicators. The weighing too should be clear.

Two conditions need to be met before a meaningful aggregate index of diverse indicators can be compiled (Innes 1990). First, there needs to be a conceptual model that which says it makes sense to add elements together. The index should correspond to an idea we can understand. For instance, the Consumer Price Index and the Gross Domestic Product are comprehensible concepts - respectively the price of a basket of goods and services and the value of production of the nation. Secondly, there needs to be a reasonable method to transform unlike things to a common scale. The economic indicators have money as a common unit of measurement. Many quality of life indices struggle to meet this condition (Rossi and Gilmartin 1980). How to combine noise nuisance (measured in decibel) with proximity to shops and public services (measured in kilometers)?

#### 8.4.10. Differentiation for purposes of performance information

The research function will benefit from a broad range of assessments. Comparisons throughout time and the confrontation with scientific standards can be a starting point for evaluation purposes. Comparison with other organizations other organizations may be beneficial for learning purposes. It may set the organization on track of better practices elsewhere. In addition, causality research can answer the question why a result is showing and disaggregating can trace results back to sections of the organization, target groups or geographical circumscriptions. This is obviously vital information for remedying insufficient performance. Less useful for research and learning are political and ideological norms and highly aggregated indices.

The internal management function will also refer to past performance in order to assess whether a result is good or not. Other organizations within the sector can be reference for output levels. Organizations outside the sector can only be relevant for comparing the management functions of the organization such as human resources management, facility management. In particular cases, scientific standards are the point of reference. Foreign organizations mostly operate in an institutional and political context that is too different as to be useful for the internal management function. In order to fulfill the internal management function, disaggregated information will be needed. For instance, the allocation of resources based on performance data requires a good cost-accounting system that is capable to provide unit costs. The causes of performance results, although not totally unrelated, are not the prime focus of this category of use. Usually, the causal assumptions will be more implicit rather than explicitly tested.

Comparing results with previous years can fulfill the accountability function. Doing better than previous times is a widespread requirement for public organizations. Comparisons with organizations within the sector are a second important norm for accountability purposes. Organizational report cards (league tables) compare schools, hospitals, universities, police units, local communities, etcetera. When the service provider is a monopolist in a country, international comparisons are used. Furthermore, scientific norms can have an accountability function. Political and ideological norms even have

accountability as a primary purpose. A report card on the management departments of organizations from different sectors appears less appealing -one prominent example being the Government Performance Project (Barrett and Greene 2005). Just like the internal management function, the causes of performance usually are more implicit than an explicit subject of research. In contrast to the research and internal management function, the accountability function requires highly aggregated information that allows for a judgment at first glance.

<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
<b>Norm setting</b>	<b>Norm setting</b>	<b>Norm setting</b>
✓ Time (+)	✓ Time (+)	✓ Time (+)
✓ Other organizations within the sector (+)	✓ Other organizations within the sector (+)	✓ Other organizations within the sector (+)
✓ Other organizations outside the sector (+)	✓ Other organizations outside the sector (+)	✓ Other organizations outside the sector (-)
✓ Foreign organizations (+)	✓ Foreign organizations (-)	✓ Foreign organizations (+)
✓ Scientific standards (+)	✓ Scientific standards (+)	✓ Scientific standards (+)
✓ Political and ideological norms (-)	✓ Political and ideological norms (-)	✓ Political and ideological norms (+)
<b>Aggregation</b>	<b>Aggregation</b>	<b>Aggregation</b>
✓ Aggregated (-)	✓ Aggregated (-)	✓ Disaggregated (-)
✓ Disaggregated (+)	✓ Disaggregated (+)	✓ Aggregated (+)
<b>Research for causality (+)</b>	<b>Research for causality (-)</b>	<b>Research for causality (-)</b>

Table 31: Differentiation of the choice of analysis technique according to the purpose of performance information (-; less useful, +; more useful).

#### 8.4.11. Step 6: reporting

The last step in the production process is the reporting of the information. The format should be apt for the target group (Rossi and Gilmartin 1980; Hendricks 1994). Reporting of performance information to top management will require other reporting format compared to reporting to media or interest groups. Different reporting formats can make performance information suitable for different target groups. The match between supply and demand in this step refers to the right reporting format for the right target group. Two questions should be answered.

*Who is consuming the information?* This is the first question. The most important target groups of performance information are represented in. The first category, the general public, is the proposed target group of many initiatives. In reality, it is hard to reach a significant part of the population. The most evident way to reach the general public is through the mass media (for instance by buying publicity).



- ✓ the general public
- ✓ mass media: newspapers, radio, television
- ✓ interest groups
- ✓ advisory boards
- ✓ international institutions
- ✓ other governments
- ✓ executive politicians
- ✓ parliament
- ✓ the board of the organization
- ✓ top management
- ✓ middle management

Table 32: Target groups of performance reporting

*What is the right format?* There are different formats for reporting performance information. gives the main options. Annual reporting for instance will be a good instrument for reporting to stakeholders and interest groups. Annual reports are for specialists. It would be unreasonable to expect a direct impact on the public in general. Oral communications will be suitable for reporting to the middle and top management. News flashes and publicity are instruments to reach the general public through the mass media.

- ✓ Annual reports and planning
- ✓ Financial documents: budget and accounts
- ✓ Specific publications in hard copy and/or on a website
- ✓ Interactive information on a website;
- ✓ oral witnesses
- ✓ news flashes
- ✓ publicity
- ✓ scorecards

Table 33: Formats of performance reporting

#### 8.4.12. Differentiation for purposes of performance information

The accountability function potentially serves the broadest set of actors. Accountability relations can be manifold. Accountability is a relationship in which an individual or agency is held to answer for performance expected by some significant “other” (Romzek and Dubnick 1998). The ultimate significant other in a democracy is the general public. The mass media are go-betweens between the organization and the general public. Interest groups in essence have the same function albeit for a particular concern. Significant others however may also be in the political system, both in the executive and the legislative system. Thirdly, supranational institutions may hold governments accountable. The Maastricht criteria and the Lisbon indicators are examples of an supranational institution (the European Union) holding the member states accountable. A broad range of reporting formats will be needed to reach this broad range of actors. Interactive information and specialized reports may be useful for specialists of interest groups. News flashes and publicity will reach a broader public. A good fit between the target group and the reporting format is required.

<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
✓ interest groups	✓ the board of the organization	✓ the general public
✓ advisory boards	✓ top management	✓ mass media:
✓ executive politicians		✓ interest groups
✓ the board of the organization		✓ executive politicians
✓ top management		✓ supra-national institutions
		✓ parliament
✓ Specific publications in hard copy and/or on a website	✓ Interim reports	✓ Annual reports and planning
✓ oral witnesses	✓ Interactive information on intranet	✓ Financial documents: budget and accounts
	✓ scorecards	✓ Specific publications in hard copy and/or on a website
		✓ Interactive information on a website
		✓ oral witnesses
		✓ news flashes
		✓ publicity
		✓ scorecards

Table 34: Differentiation of the reporting format according to the purpose of performance information.

### 8.4.13. Throughout the process: quality assurance

Quality assurance of performance information is a crucial matter. Quality needs to be guaranteed for two reasons. First, bad information may be used, since information is intended to be used. Consequently, it may lead to wrong decisions and actions. Users of information (decision makers, politicians, media) often have a limited capacity (lack of time and/or competences) to assess quality of information. Therefore, information is trusted. Secondly, bad information may not be used. If the users of information learn about the weaknesses of information, they probably will disregard it. It will be hard to regain trust in information.

Quality concerns the whole production process of performance information. The previous steps in the process are necessary conditions for the next steps. Indicator development can only be done properly when the subject of measurement is defined. Focused data collection can only be done when the indicators are available. Meaningful analyses are only possible with qualitative data. Reporting for the target groups is only possible with the right analysis. Thus, when quality of performance information is a problem, this may be due to diverse steps in the measurement process.

The organization of quality assurance of performance information ideally parallels the control pyramid used by auditors (Sterck, Scheers and Bouckaert 2005). The first level is the internal control system of the organization itself. The second level is the internal audit that controls the control processes and assesses the risks. The internal audit reports to the management of the organization. Thirdly, the external audit reviews the quality independently from the organization. For what financial information is

concerned, this system is well established. Non-financial information is seldom included in the audit systems (Wholey 1999).

Quality is not a mere technical matter. There are three aspects to quality (Bouckaert 1993).

- First, quality implies the functionality of the measurement system. Measurement should be fit for use. There are two gradations of non-conformity to the functionality requirement: non-functionality and dysfunctionality. First, non-functionality implies that the information is disregarded. The measurement efforts and costs are in vain. Secondly, dysfunctionality implies that there are negative effects due to the measurement. The organization in that case is worse off than before (infra).

- Secondly, quality implies indicators that are valid and reliable. First, a reliable indicator yields the same values for repeated measurements of the same object. Reliability answers the question whether measurement is good. Secondly, a valid indicator measures what one intends to measure. It is about measurement of the right things. Does the measurement reflect the measured reality well?

- The third quality dimension is legitimacy of a measurement system. In an ideal scenario, all organization members support the measurement system. Manipulation and gaming with performance information are less likely when the ownership is high. Only when unobtrusive indicators exist, ownership may be less vital for the measurement effort. Legitimacy may be internal or external to the organization. The distinction refers to respectively internal actors and external actors.

The functionality criterion should be assessed first. Performance measurement systems that are functional or even dysfunctional should be abolished or redesigned. The other two criteria may be applied more variably accordingly to the function of the information for the organization.

#### 8.4.14. Differentiation for purposes of performance information

Before we make the assessment for different functions of performance information, two remarks need to be made. First, we use the categories 'moderately important', 'highly important' and 'critical'. This is because a certain level of quality is required in all cases. It would be wrong to state that a particular quality characteristic is not important for a particular use. Yet, there is still a variation in importance. It may be argued that quality is not an absolute qualification. The marginal costs and benefits of more quality need to be taken into account<sup>76</sup>. Secondly, the quality dimensions may influence each other as

---

<sup>76</sup> Hatry (2002) for instance argues that pressures from the professional community have overstressed the need for high levels of precision and

independent and dependent variables. For instance, a high legitimacy within an organization may lead to high reliability because employees are mindful with regard to registrations. In addition, high validity may lead to high legitimacy because people feel the right things are measured.

The research function poses the least strict conditions on validity and reliability. The performance information will be complemented by other information sources such as individual experiences of employees. Crucial however is the legitimacy of the measurement effort. Without the belief that measurement may allow for evaluating, learning and improving, the research function will not be fulfilled. In addition, the flexibility that is required to interpret the performance results will be absent when legitimacy is low. Therefore, legitimacy is crucial. When policy results need to be assessed, external legitimacy is of prime interest. When management and operational results are concerned, the prime focus is on internal legitimacy.

Secondly, quality requirements that are mainly important for the internal management function (to allocate, to motivate, to control) are validity and reliability. When controls, rewards, budgets and personnel are distributed in the organization, validity and reliability are critical, in particular when the degrees of freedom are heavily restricted by the use of performance information and when there is a mechanic 1:1 relation between performance results and management decisions. Internal legitimacy is of second order importance. It is mainly a manager's responsibility to control and to allocate resources. In particular, external legitimacy of the measurement system is moderately important.

The external accountability purposes require high validity and reliability, because the consequences of bad or good results for the organization may be so high. The organization needs to be convinced that they are judged fairly. Obviously, external legitimacy is also critical. Internal legitimacy may be of second order importance. When the information will be used for instance for performance contracts, the legitimacy in the whole organization may be less important than the reliability and validity of the information. Report cards often are imposed on organizations. Internally, the legitimacy of a report card may be low. This is probably not problematic as long as the validity of the results is good (a fair judgment) and as long as low ownership does not affect reliability because of gaming.

---

response rates. This has discouraged practitioners. Hatry states that the operational principle should be that it is better to be roughly right than to be precisely ignorant. This is under the condition that no death-or-life issues or large amounts of money are involved.

<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
✓ Validity *	✓ validity ***	✓ validity ***
✓ reliability *	✓ reliability ***	✓ reliability ***
✓ internal legitimacy ***	✓ internal legitimacy **	✓ internal legitimacy *
✓ external legitimacy***	✓ external legitimacy*	✓ external legitimacy***

Table 35: Differentiation of the quality dimensions according to the purpose of performance information. (\* moderately important; \*\* highly important; \*\*\* critical)

## 8.5. Conclusion

Table 36 summarizes the relation between the measurement process and the uses of performance information.

	<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
Implementation mode	bottom-up incremental	rather top-down single package	top-down decremental
Criteria for the definition of the measurement object	efficiency of measurement <ul style="list-style-type: none"> <li>✓ indications of problems</li> </ul> developmental perspective <ul style="list-style-type: none"> <li>✓ added value</li> <li>✓ feasibility</li> <li>✓ distribution</li> </ul>	efficiency of measurement <ul style="list-style-type: none"> <li>✓ financial coverage</li> <li>✓ personnel coverage</li> </ul> developmental perspective <ul style="list-style-type: none"> <li>✓ distribution</li> </ul>	efficiency of measurement <ul style="list-style-type: none"> <li>✓ societal visibility</li> </ul> predetermined
Choice of indicators	<ul style="list-style-type: none"> <li>✓ the whole 'production chain of the organization'</li> <li>✓ room for interpretation</li> </ul>	<ul style="list-style-type: none"> <li>✓ mainly output and efficiency</li> <li>✓ moderate room for interpretation</li> </ul>	<ul style="list-style-type: none"> <li>✓ optimally outcome, with output as a second best</li> <li>✓ Limited room for interpretation</li> </ul>
Method of data collection (+ useful, - less useful)	<ul style="list-style-type: none"> <li>✓ Existing registrations (+)</li> <li>✓ Additional registrations (+)</li> <li>✓ Surveys (+)</li> <li>✓ Self-assessments (+)</li> <li>✓ Technical msmt (+)</li> <li>✓ External observers (-)</li> <li>✓ Other organizations (+)</li> <li>✓ Statistical, international, and research institutions (+)</li> </ul>	<ul style="list-style-type: none"> <li>✓ Existing registrations (+)</li> <li>✓ Additional registrations (+)</li> <li>✓ Surveys (-)</li> <li>✓ Self-assessments (-)</li> <li>✓ Technical msmt (+)</li> <li>✓ External observers (+)</li> <li>✓ Other organizations (-)</li> <li>✓ Statistical, international, and research institutions (-)</li> </ul>	<ul style="list-style-type: none"> <li>✓ Existing registrations (+)</li> <li>✓ Additional registrations (+)</li> <li>✓ Surveys (+)</li> <li>✓ Self-assessments (-)</li> <li>✓ Technical msmt (+)</li> <li>✓ External observers (+)</li> <li>✓ Other organizations (-)</li> <li>✓ Statistical, international, and research institutions (-)</li> </ul>

	<i>Research function</i>	<i>Internal management function</i>	<i>Accountability function</i>
Method of data analysis (+ useful, - less useful)	<b>Norm setting</b> ✓ Time (+) ✓ Other organizations within the sector (+) ✓ Other organizations outside the sector (+) ✓ Foreign organizations (+) ✓ Scientific standards (+) ✓ Political and ideological norms (-) <b>Aggregation</b> ✓ Aggregated (-) ✓ Disaggregated (+) <b>Research for causality (+)</b>	<b>Norm setting</b> ✓ Time (+) ✓ Other organizations within the sector (+) ✓ Other organizations outside the sector (+) ✓ Foreign organizations (-) ✓ Scientific standards (+) ✓ Political and ideological norms (-) <b>Aggregation</b> ✓ Aggregated (-) ✓ Disaggregated (+) <b>Research for causality (-)</b>	<b>Norm setting</b> ✓ Time (+) ✓ Other organizations within the sector (+) ✓ Other organizations outside the sector (-) ✓ Foreign organizations (+) ✓ Scientific standards (+) ✓ Political and ideological norms (+) <b>Aggregation</b> ✓ Disaggregated (-) ✓ Aggregated (+) <b>Research for causality (-)</b>
Reporting formats	✓ interest groups ✓ advisory boards ✓ executive politicians ✓ parliament ✓ the board of the organization ✓ top management  ✓ Specific publications in hard copy and/or on a website ✓ oral witnesses	✓ the board of the organization ✓ top management  ✓ Annual reports ✓ Financial documents: budget and accounts ✓ Interactive information on intranet; ✓ scorecards	✓ the general public ✓ mass media: ✓ interest groups ✓ executive politicians ✓ supra-national institutions ✓ parliament  ✓ Annual reports ✓ Financial documents: budget and accounts ✓ Specific publications in hard copy and/or on a website ✓ Interactive information on a website ✓ oral witnesses ✓ news flashes ✓ publicity ✓ scorecards
Quality criteria * important ** highly important *** critical	✓ validity * ✓ reliability * ✓ internal legitimacy *** ✓ external legitimacy***	✓ validity *** ✓ reliability *** ✓ internal legitimacy ** ✓ external legitimacy*	✓ validity *** ✓ reliability *** ✓ internal legitimacy * ✓ external legitimacy***

Table 36: Use related to design: summary table

Degrees of freedom are restricted along with the underlying function of the use of performance information. Accountability purposes heavily restrict the operational freedom of organizations. Internal management purposes restrict the internal freedom of subordinates of the managers, but do not affect the managers themselves. The research function has learning and improving as a purpose. Therefore, it does not restrict degrees of freedom. Performance information in this case is an enabler.

Performance information often adopts the characteristics of regulation (Christensen and Laegreid 2005). If accountability is defined outside organization, then the performance information has a regulatory function. If the organization itself determines the subject (accountable for what) and the target group (accountable to whom) of an accountability relation, then the performance information does not have a regulatory function. Christensen and Laegreid point out that 'one specific feature of regulation is that it is an external form of control of formal organizations (p5)'. Performance measurement systems may be forms of regulatory control. The definition of a measurement system also defines the standards of good policy and management. Both performance contracts (Christensen and Laegreid 2005) and report cards (Gormley and Weimer 1999) may play the role of being authority tools to require or prohibit organizational behaviors (Schneider and Ingram 1990)

The multitude of design parameters shows that a good fit between supply and demand of performance information is only partly a matter of the right quantity of performance information. The main challenge is to provide the right qualities with the right quality. This chapter gives an indication of the differentiation of the choices in the measurement process for the main functions of performance information.

The options in the measurement system can be incompatible or complementary. In the former case, one option excludes another. For instance, performance information for learning purposes requires considerable room for interpretation in the analysis. Performance information for accountability usually does not allow for this flexibility. Another example is the variation in the implementation mode of a measurement system. A top-down and a bottom-up approach are to a large extent mutually exclusive. In the case of complementary choices, the options can be cumulatively combined. This leads to a more extensive measurement system. Examples are the reporting formats, the choice of the indicators and the methods of data collection and analysis. Also, the performance measurement system should score high on all quality criteria in order to serve multiple functions.

Because of the incompatibilities, and because of the cost of an extensive measurement system, multifunctional measurement systems are not obvious. Performance information that has the right qualities for all three functions is exceptional. A measurement system that can provide information for all three functions too is rather uncommon. Regularly, different separate measurement processes run parallel in organizations - e.g. one in the quality department, one in the budget department, one in the internal think tanks, etcetera. A better integration of measurement processes may yield economies of scale. However, the incompatibilities in measurement design may be an obstacle. How can this problem be mitigated?

A first way is to primarily integrate measurement processes that serve the same functions. Measurement for evaluating, learning and improving internal processes on the one hand and external outcomes other, may be integrated in a single process. Presumably, the combined effort will show

better results than the sum of the parts. In the same way, measurement systems for control and allocating resources (typically in the personnel department and financial department of the organization) may be integrated.

A second option is to integrate primarily the step of the data collection. This is usually the most costly step. It is costly in financial terms. New data collection, for instance through surveying is usually not cheap. There are also the hidden costs of administrative overhead to consider. The time that is invested in (new) registrations -and is diverted from front line service delivery- may quickly add up. A good registration system may feed several methods for analysis and several reporting formats with data. According to the function, this system of data registration can be complemented by other sources of data.

Thirdly, the problem can be overcome by investment. By building an extensive measurement system with a good quality assurance system, the different functions of performance information may be fulfilled. Obviously, for many organizations, this is not an option. In most countries, public budgets are under pressure due to increasing global competition (Pollitt and Bouckaert 2004). Investment in information is regularly seen as an overhead costs that comes at the expense of investment in the core business. Or worse, it is regarded as the toy box of top-managers. This makes it hard to argue with decision-makers who have to decide on extra budgets.

The diverse uses of performance information require an adapted design of the measurement system. The adaptation involves all aspects of a measurement system, and not only the selection of the performance indicators. This integral focus is required to obtain performance information with the right qualities



## 9. What are the effects of performance measurement?

### SUMMARY OUTLINE

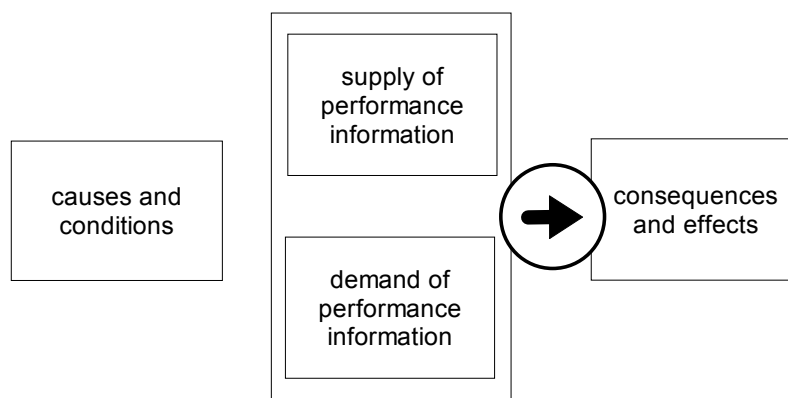
**Issue:** Measurement of performance is not neutral. Since the Hawthorne experiments, we know that the act of measurement in itself is influencing performance. By measuring performance of organizations, the organizational behavior can also be affected. Several authors addressed this issue. This resulted in laundry lists of potential unanticipated effects of performance measurement. This section will first describe the most important effects from the literature. Next, we will look search for a common denominator. Subsequently, we will establish which are the empirically most important effects. We then will examine the hypothesis that the effects are a consequence of a specific profile of use. This viewpoint challenges the tendency of effect-studies to view performance measurement as a one-dimensional concept.

### Research questions:

- What are the effects of performance measurement in the literature?
- What is the common denominator in these effects?
- What are empirically the most important effects?
- Which effects arise from which use?

**Methodology:** literature study, survey data and semi structured interviews for 12 sections (N=12)

### Graphical representation



## 9.1. Introduction

A thermometer has no impact on the temperature. Measurement is unobtrusive. Performance measurement in organizations - which are social realities - is fundamentally dissimilar. Elton Mayo (1933) showed with the Hawthorne Experiments that the mere fact of being observed does influence behavior. People react to being measured. Measurement in organizations is never neutral (Ridgeway 1956). Performance measurement too is not neutral to the organization or the policy sector in which it is introduced. Performance measurement is believed to be the cure for much malfunctioning in organizations and policy sectors. Yet, if performance measurement is the medicine, the information leaflet looks at least disturbing. The organization may contract hypertrophy or atrophy, myopia, tunnel vision and ossification (Smith 1988; Bouckaert and Balk 1991; Bouckaert 1995; Smith 1995). The effects of measurement are well documented. Yet, it remains a research interest for Public Administration. The U.K. Economic and Social Research Council launched a research program titled 'The Public Services: Quality, Performance and Delivery' as recent as January 2005<sup>77</sup>. One of the main topics are the effects of popular reform measures like incentive pay, targets, transparency.

This section studies the effects of performance measurement. We will argue that the common denominator of the laundry lists of effects is the principle of goal displacement (Merton 1957; Blau and Scott 1962). The effects then will be linked to a particular use of performance information. The use of performance information is the independent variable. Rather than regarding performance measurement as a one-dimensional concept, we hypothesize that particular effects are causes of particular uses of performance information.

First, we will discuss the literature about effects of performance measurement. We will look at measurement issues on the one hand and output issues on the other. Secondly, the concept of goal displacement is introduced as a common denominator for the behavioral issues. Thirdly, we put forward some hypotheses about both use and effects. Fourthly, the data and the results are described. Next, we will discuss the distinction between the motive and the opportunity for skewing performance information.

---

<sup>77</sup> <http://www.publicservices.ac.uk/> (accessed 15/11/2005)

## 9.2. A multitude of effects

A multitude of effects of performance measurement has been described (Poister 2003). Generally, the effects are caused by either a manipulation of the measurement process or a manipulation of the organizational output (Figure 36)<sup>78</sup>. The first set of effects mainly affects the measurement process. Measurement is skewed due to these dysfunctions. Yet, the output of the organization is not necessarily affected. This is represented by (1) on Figure 36. For instance, in a case where measurement is pure window dressing, bad measurement may not impact the day-to-day operations of the organizations. Secondly, some dysfunctions do alter the daily operations of the organization. The practices of organizations are changed. Different quantities and/or qualities of output may be pursued. These dysfunctions may materialize even with a perfect measurement system - (3) on Figure 36. However, with an imperfect measurement system, the chances of operational dysfunctions are higher - (2) on Figure 36. Flawed measurement is a multiplier for distorted practices. Flawed measurement results may become a self-fulfilling prophecy.

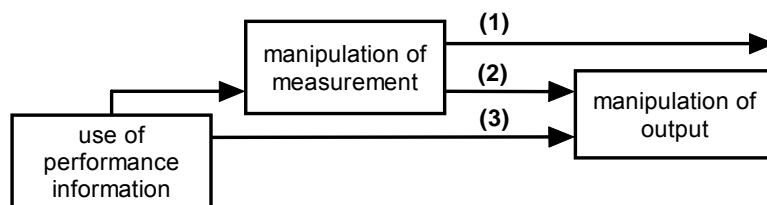


Figure 36: Effects of the use of performance information

### 9.2.1. Manipulation of measurement

Manipulation of measurement -intentionally or not- comes in many guises. There are at least seven ways of manipulating the measurement process (Smith 1988; Bouckaert and Balk 1991, Smith 1995).

First, measurement may be convex or concave. The measured value does not correspond with the real value. Convexity gives us the perception of a higher than real value. Concave measures show a lower than real value. Examples of both can be found in bibliometrics, which attempt to measure scientific impact through citation analysis (Garfield and Welljams-Dorof 1992). On the one hand, the

---

<sup>78</sup> In this study, we only looked at the effects of measurement in organizations that do have measurement systems. Bouckaert and Balk (1991) add the Pangloss disease ("measurement is not needed, because nothing goes wrong"), impossibility disease ("it is necessary but not possible to measure performance") and hypochondria ("measurement is useless because the public sector can never obtain the results of the private sector"). These predispositions towards measurement are not studied here.

phenomenon of citation circles leads to convexity. It brings about an overestimation of the impact those authors that cite each other. On the other hand, the obliteration phenomenon leads to concavity. This phenomenon refers to a process in which breakthrough advances -for example Einstein's theory of relativity - are cited less frequently over time. These landmark discoveries are incorporated into the generally accepted body of scientific knowledge. They are assumed, and therefore no longer cited. The measured value underestimates the real impact

Secondly, Bouckaert and Balk (1991) refer to the Mandelbrot disease as failing measurement. Mandelbrot, a mathematician, measured the length of the British coastline (1967). He shows that by refining the measurement scale, the length of the coastline approaches infinity instead of a limit. At any resolution, more inlets and peninsulas are visible that were not visible before. Thus ,by looking at increasingly fine resolutions, more and more lengths are approximated, and the total estimate of length appears to increase without bound. This process also takes place in performance measurement. More measurement points may lead to higher values because things are observed that were not seen before. An example is the number of violations of human rights reported by Amnesty International. This may be because of a real worsening of the situation. Yet, a higher number may also be caused by the establishment of a higher number of observatories. The more one looks, the more one sees.

Thirdly, the number of indicators in a set of indicators often risks inflating. This process is often termed as the 'mushrooming' of indicator sets. Too much indicators may indeed be problematic. The users of the information cannot see the wood for the trees anymore. Yet, the addition of increasingly more indicators is often a remedy of at least two other performance measurement effects: cream skimming (selecting the intake) and tunnel vision (focusing on the measured activities only). There is probably no optimal number of indicators. Organizations need to find this out through experience. The Audit Commission in the United Kingdom for instance gradually reduced the number of indicators on local government from 242 in 2000 to 166 in 2002 (Bouckaert, Depeuter and Van Dooren 2003).

Fourthly, measurement systems may get polluted (Bouckaert 1995). This mainly refers to the terminology. Different people interpret the concepts and definitions (slightly or substantially) differently. For some, performance refers to output, for others, performance is outcome and for still others it means both. Pollution may appear a trivial matter. Usually, there is no deeply rooted antagonism. Yet, and maybe because of its triviality, agreement is seldom sought. As a consequence, people talk at cross-purposes and the effectiveness of performance measurement erodes significantly.

Fifthly, performance information may be manipulated by aggregating or disaggregating data (Winston 1993; Perrin 1998). Lesser performance may be obscured by more or less aggregate indicators. Separate indicators can be combined in composite indicators. Composite indicators have the benefit of simplicity. Decision makers with limited time or the general public with limited insight into complex

policy matters are helped with a universal assessment of performance. Yet, by choosing and weighing the measures, organizations may hide problematic aspects of their performance. It may also happen the other way round. An organization may look for more detail until the performance is satisfactory. For example, the age cohort may be defined in such a way that the policies appear effective.

Sixthly, misrepresentation is the deliberate manipulation of data –ranging from creative accounting to fraud- so that reported behavior differs from actual behavior (Smith 1995). In the United Kingdom, near impossible ambulance response-times of less than a minute were reported. One case even reported a outright impossible response time of less than zero seconds (UK House of Commons, Public Administration Select Committee 2003). Clearly, reality is misrepresented. Financial information systems combat misrepresentation by installing extensive internal control systems, supplemented by internal and external audit systems (Raaum and Morgan 2001; Sterck, Scheers and Bouckaert 2005). For non-financial information, there is usually not such an extensive control structure. The cost may seem too high while the benefits are intangible and remote. Misrepresentation thus has prevented in a different way. Some authors propose a more heavy reliance on trust-based systems (Power 1997; Grizzle 2002).

Seventhly, misinterpretations are the incorrect inferences about performance brought about by the difficulty of accounting for the full range of potential influences on a performance measurement (Smith 1995). This is particularly the case for outcome measures. The ultimate results of government output in society are often only visible in the long term. Moreover, they are seldom the merit of one single agency. Low unemployment figures for instance may be the result of a good training program. Yet, the economic climate is without doubt also an important determinant. Moreover, tough inspection on moonlighting may add to the overall result of declining unemployment.

### 9.2.2. Manipulation of output

Manipulation of output is a second category of responses to the use of performance information. Unlike the manipulation of measurement, these responses alter the output of the organization. Measurement is not (necessarily) skewed, but the underlying reality is altered. In the remainder of this section, we will primarily deal with this category. We discuss ten effects.

First, organizations that measure performance may contract hypertrophy (Bouckaert and Balk 1991). In medicine, hypertrophy is an enlargement or overgrowth of an organ due to an increase in the size of its constituent cells. Similarly, in public administration, performance measurement may cause a specific output to boom because it is measured. This is especially the case when costs per unit calculations are being made. When fixed costs are considerable, it usually is easier to reduce cost per unit by increasing output rather than decreasing input. As a result, the total output inflates (Dawson and Street 2000).

Secondly, atrophy is the opposite since outputs are reduced rather than inflated (Bouckaert and Balk 1991). For instance, quality of output may decrease because of a crude quantitative measurement. Heinrich (1999) for instance observed that a predominance of cost-per-placement considerations in a job-training program had a negative impact on service quality. This effect was also described by Berliner (1956) in the context of the Soviet production targets. The use of monthly production quota led to 'storming' at the end of the month. Repairs and maintenance were postponed to the next month which led to a new rush at the end of the next month. The main loss occurred because suppliers and consumers were forced into a fluctuating rate of operations. A third example comes from a British hospital, where waiting-time targets led to cancellations, long waiting times before appointments could be made, and a lack of follow up visits. Measurement leads to less rather than more output (UK House of Commons, Public Administration Committee 2003).

Both hypertrophy and atrophy are instances of what Smith (1995) called 'measure fixation'. This is the pursuit of success as measured rather than as intended. Smith (1995) gives the example of the 'hello' nurse who had to make the first contact with patients in order to meet a five-minute waiting time requirement. Although the target is being met, this clearly is not an improvement of service delivery. Another example are university rankings. Prestige rather than quality of the programs is the main factor driving rankings (Frederickson 2001; Gormley and Weimer 1999). To obtain good rankings, universities may primarily try to improve prestige and not program quality<sup>79</sup>.

A third dysfunction is myopia (Bouckaert 1995). The long-term view is excluded by an over-fixation on the short-term goals. This effect can be found in the private sector too. Companies are often urged on constantly improving quarterly results. Herewith, the sustainability of profits in the long run may be neglected. In the public sector, myopia usually favors curative services above preventive services - for example crime solving rather than crime prevention. Prevention is an example of an activity with per definition intangible results. In fact, the results are what did not happen. Another example can be found in Blau's work (1955). In a court the target of eight cases a month per person leads to the postponing of difficult cases in favor of easy cases. The initial prioritization criterion, relative urgency, was replaced by another criterion, relative ease of processing.

Thirdly, organizations that measure performance may suffer from tunnel vision. In this case, organizations only pay attention to those activities that are being measured. This means that important

---

<sup>79</sup> Frederickson (2001) mentions several strategies such as attracting faculty that publishes in respected journals, nurturing a doctoral program with a good outplacement in recognized education programs, participate visibly in associations, attract an important journal editorship, develop a well-oiled alumni organization, develop visibly one or two specializations, have at least one faculty member publishing in the popular press, and have at least one "grantsman-rainmaker" in the faculty.

unmeasured areas are neglected (Smith 1995). For example, intangibles such as training and advice may be ignored. Blau (1955) gives an example of tunnel vision in an employment agency. Interviewers were motivated to complete as much interviews as possible. By doing this, they did not pay sufficient attention to other activities such as locating new jobs. A tunnel vision may be avoided by including more indicators. However, by adding indicators the risk of mushrooming is increasing too..

Fifthly, sub-optimization refers to the pursuit of local organizational objectives at the expense of supra organizational objectives (Hood 1974; Bouckaert and Balk 1991; Smith 1993; Perrin 1998). The optimal result for a single organization may not always be optimal in a broader context. This is especially immanent when the production of an outcome is the responsibility of a sequence of several actors. Security is such a outcome of a chain of outputs. First, there is prevention. For instance, public places need to be well lit at night. People in shopping areas need to be alerted for pickpockets. Next, the police needs to patrol and make arrests. The public prosecutor has to institute legal action. The courts have to pass judgments. Finally, the prisons have to detain convicts. Ideally, the social services run programs to reintegrate detainees into society. The optimal result is as much safety as possible. A police force may score well on an indicator such as the number of arrests by arresting more quickly, including the lesser offences. This may be optimal for the police force, but suboptimal for the ultimate outcome. The limited capacity of the judicial system may become strained, and the more serious crimes may remain unsolved.

The sixth reaction is cream skimming (or cherry picking) (Behn and Kant 1999; Grizzle 2002). Organizations may be tempted to select the intake. Easy cases and clients are processed while the more difficult cases are redirected. Job training programs for instance could select the unemployed that are most likely to find a job (Anderson, Burkhauser and Raymond 1993; Heckman, Heinrich and Smith 1997). This may be economically efficient, but it usually contrasts with the political goals of public programs. This effect is conditional. Demand for the service needs to be higher than supply and the selection decision needs to be made by the organization itself.

Seventhly, complacency is the lack of ambition brought about by adequate comparative performance. Many organizations may want to stay securely in the pack. The French phrase *pour vivre heureux, vivons caché* (in order to live happily, we need to live concealed) is emblematic for this position. The reasoning is that both top and low performers may lose budgets. The result of complacency is stagnation. Overall performance will tend towards the average. This effect may lead to the performance paradox (Meyer and Gupta 1994; Van Thiel and Leeuw 2002). Throughout time, the indicator does not discriminate between bad and good performers because organizations adapt to each other.

Eighthly, gaming refers to altering behavior and output for the strategic reasons (Smith 1995). This is particularly the case in principal agent relations, where targets are set externally. Organizations may

deliberately lower performance in order to avoid higher targets for the next year. Principals are tempted to increase the standard for superior performers. Agents foresee this possibility and may be hesitant to supply bigger effort in order to avoid increases in their targets. Target inflation is also termed the ratchet effect. Fear of the ratchet effect annuls the incentives that a measurement system is believed to introduce (Courty and Marschke 2003).

Ninthly, ossification in medicine is the formation of bone in the body. Analogously, excessively rigid measurement system may lead to organizational paralysis (Smith 1995). For instance, too detailed time registration systems may inhibit experimentation. Innovation inherently requires some tolerance towards failure. When time registration systems make every failure visible and the cost computable, innovation may be inhibited.

Finally, measurement and target setting may lead to polarization of good and bad cases. Some services or cases may be hopeless in the light of the target. Rather than investing more resources to get these services to an acceptable level, organizations may decide to give up these services and to invest in the marginal cases. This may be termed as the total loss effect, since some cases are virtually thrown away. An example may be found in railroad companies. The indicator typically is the percentage of trains arriving on time. A railroad company would rather have one train being much too late or even cancelled than many trains being a little late (De Morgen 20 September 2005).

### **9.3. Goal displacement; for good and for bad**

We identified seventeen responses to performance measurement that are usually very familiar to both practitioners and researchers of performance measurement. Although the effects may be very recognizable, the sheer length of the lists (which in all probability is still incomplete<sup>80</sup>) calls for a unifying theoretical foundation. In this text, we will focus on ten effects where the output quantity or quality is altered. Mostly, these are the effects with the highest impact, since real service delivery is altered and not only its portrayal. The concept of goal displacement may serve as common denominator for the list of responses to performance measurement (Bohte and Meier 2000).

Goal displacement is the process whereby “*an instrumental value becomes a terminal value*” (Merton 1968). The means become ends. Performance indicators are means for better planning, managing and evaluating organizations and policies. Yet, in all ten cases, the measures are becoming ends at the expense of terminal values such as better service delivery and policies. These responses are

---

<sup>80</sup> Gerald Caiden identified 132 bureaupathologies from the literature (1991). These certainly do not all arise from measurement. Yet, it gives an indication of the multitude administrative dysfunctions.



dysfunctional strategies to score well on the indicators. Creaming for instance excludes those people that may benefit the most of a program (Perrin 1998). Complacency with performance and ossification lead to a loss of innovation. In all ten cases, goal displacement is seen as a negative. The change in organizational output is considered harmful. The result of goal displacement is the development of a measurement culture, at the expense of performance culture (U.K. House of Commons – Public Administration Select Committee 2003).

However, the proponents of performance measurement are also concerned with goal displacement. Performance measurement has to bring about goals displacement (1), or has to avoid it (2).

First, in some cases goal displacement is simply the aim of performance measurement (1). The introduction of performance indicators is done precisely to alter organizational output. This course of action is a central assumption in principal agent relations. This literature states that monitoring is one of the main transaction costs in aligning the divergent objectives of principals and agents (Alchian and Demsetz 1972). Performance contracts between central government and autonomous agencies may serve as a tool in making agencies work towards the goals of government (Verhoest 2002). Another example is national indicator initiatives such as the Lisbon criteria of the European Union. The Lisbon indicators are a set of indicators and targets on innovation, economic and social performance agreed upon by the member states of the European Union. The idea is to align national goals with the European program. This approach is called ‘the open method of coordination’ (European Commission 2000; 2001a; Tucker 2003)<sup>81</sup>. In these cases, the instrumental value, the performance indicators, are means to force organizations or governments in the direction of particular terminal values. These terminal values may differ from the initial values of the organization.

Secondly, in other cases performance measurement has to avoid goal displacement (2). Hatry (1999) for instance states that measuring performance should increase the ability of managers to get the job done with the resources they have. Managers need to know the running score in order to keep the organization on track towards their goal. By measuring progress towards goals, the organization avoids pursuing the erroneous goals. Again, the instrumental value, the performance indicators, is a means to keep the organization on track of the terminal value.

---

<sup>81</sup> The Lisbon indicators had to map progress towards the goal of making the European economy the most dynamic one in the world by 2010. Each year, the European Commission reported on the progress towards this goal (European Commission 2000; 2001b; 2002;2003). Recently, it became clear that the target will be missed (European Commission 2004). After this mid-term review, the 2010 time horizon was replaced by a three year cycle of planning and evaluation (European Commission 2005).

Goal displacement thus is either functional or dysfunctional. Essentially, the original formulation of the concept by Merton was about rules in bureaucratic organizations (1968). Conformance to regulation is not seen as a measure for a specific purpose, but for becomes an immediate value in life-organization of the bureaucrat (p. 253). We would expect that goal displacement, and the different strategies to get there, is more likely with performance systems that have the qualities of regulation. The hypothesis claims the more a measurement system infringes on the discretion of an organization, the more goal displacement processes will show.

The reduction of discretion however is assumed a result of the way information is used and not of the mere provision of performance information. Thus a more precise hypothesis is that diverse uses reduce discretion of organization to a dissimilar extent, and therefore bring about goal displacement strategies.

#### **9.4. Central hypothesis: use determines effects.**

Before we elaborate on the central hypothesis, we briefly repeat the three categories of use we defined above (p. 160). Three functions are identified: research, internal management and accountability (The Working Party on Performance Monitoring in the Public Services of the Royal Statistical Society 2005).

1. First, performance information may be collected in order to find out what works? This is the research and learning function. The main rationale is indeed learning. The rise of evidence based policy initiatives may be seen in this light (Davies, Nutley, and Smith 2003; Boaz and Nutley 2003). People are intrinsically motivated to perform well and to seek responsibility. This also motivates them learn how to do the job. How can policy or management be improved? Policy planning and evaluation, strategic planning and evaluation and business process re-engineering are examples of policy and management tools with a primarily research orientation. Performance information can be used for process evaluation and outcome evaluation, which envisages respectively service improvement as well as policy improvement (Scheirer 1994). The use of performance information for research and learning has the lowest impact on the degrees of freedom of the organization.

2. Secondly, the internal management function of performance information is about identifying and sanctioning well or underperforming institutions or public servants (to control, to motivate) and about allocating resources (to budget). The purpose is not to learn from mistakes, but to agree upon targets. The use of performance information for management tools such as internal performance contracts, performance based pay, performance budgeting and performance mandates fall in this category (Bovaird and Loffler 2003; Kamensky and Morales 2005). The use for internal management may have a significant impact on the degrees of freedom within the organization.

3. The third purpose is accountability. The main proposition is that the public sector should be accountable to the citizens/taxpayers and politicians (Bolton 2003). Therefore, performance of public bureaucracies should be disclosed to the public. Two categories of external pressure can be identified. First, pressure is assumed by showing results to the general public. In case of a monopoly, the potential criticism of the public is expected to exert enough pressure on the organization. In case of (quasi-) markets, rankings are supposed to provoke market pressures. Typically, league tables, citizen charters and upgraded annual reporting are examples of this accountability role (Bowerman 1995; Gormley and Weimer 1999). Secondly, pressure is instituted by the political system. Performance contracts with agencies are an example. These contracts give autonomy to agencies provided the agency commits itself to output or outcome targets. The use for accountability usually has a high impact on the degrees of freedom of the organization.

Figure 37 represents the central hypothesis. Utilization of performance information determines effects. The more performance information is used for accountability, management and research, the more effects will show. The function starts at the origin. No utilization will not cause changes in organizational outputs and behavior. Yet, this may not always be the case. Organizations that are not measuring may experience the effects of measurement. This may occur when peer organizations are measuring and they are subjected to some pressure to align to their peers. In that case, the function would not start in the origin

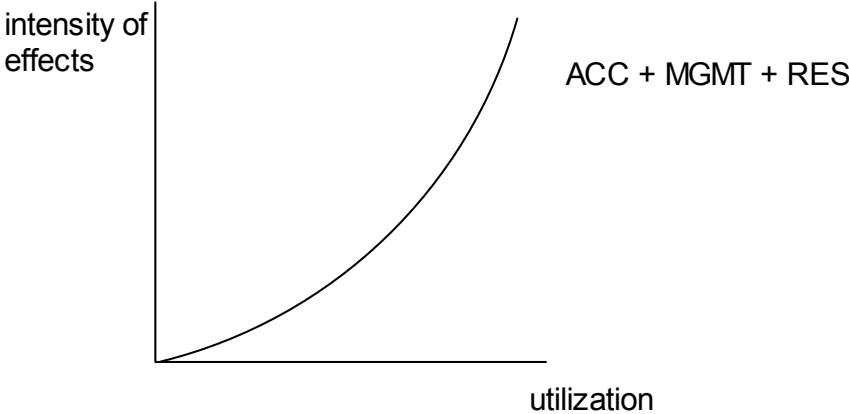


Figure 37: Central hypothesis of measurement effects: utilization determines effects

The function is hypothesized to be exponential. This is because it adds up the three functions for the different categories of use. The hypothesis is that with increased utilization, the effects will also increase. Yet, the alteration (delta) differs for categories for use. Use for research and learning will cause fewer effects than use for management or accountability.

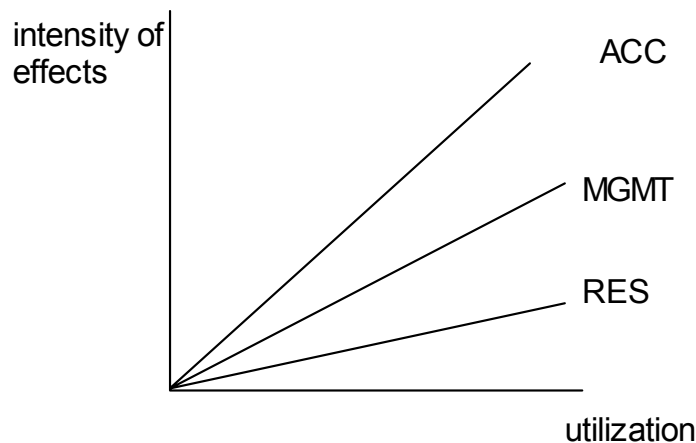


Figure 38: Central hypothesis of measurement effects: breakdown for categories of use

This then sets the following hypotheses. First, high use will lead to high effects. Given our methodology (interview-setting), this hypothesis states that organizations that are measuring more will perceive more effects of measurement. Second, high utilization for accountability will lead to high effects, no matter what the utilization for research or management is. The different alteration (delta) of the three functions implies that effects of accountability overtake effects of other uses in case of high use for accountability. Secondly, high utilization on management, but low on accountability will lead to moderate effects, no matter what the utilization for research is. Thirdly, high use for research will lead to low effects, under the condition that use for management and accountability is low.

*H1. high use will lead to high effects: organizations that are measuring more will attach a higher sense of reality to the effects.*

*H2 high use for accountability will lead to the highest effects, no matter what the utilization for research or management is.*

*H3 high use on management, but low on accountability will lead to moderate effects, no matter what the utilization for research is.*

*H4 high use for research will lead to low effects, under the condition that use for management and accountability is low.*

The underlying assumption about the relation between the categories of use is that high impact use drives out low impact use. Impact in this case is the impact on the degrees of freedom of the organization. This is a common observation in social psychology. Intrinsic motivation is likely to be undermined by extrinsic motivation (Deci 1975; Lepper and Greene 1978). Research and learning are

built upon intrinsic motivation. Internal management control and accountability are based on extrinsic motivation. Therefore, the latter uses may override the former.

### 9.5. Data

Figure 39 gives an oversight of the research process. The emphasis lies on twelve semi structured interviews. We chose face-to-face interviews for two reasons. First, it complements the survey data. Secondly, and more importantly, the research issue - the effects of performance information - is not suited for survey or document analysis. Perverse and negative effects in particular are prone to social desirable answers. Therefore, a survey would yield deceptive results. Equally, these perverse effects are not found in documents, since organizations will precisely attempt to avoid publication of these mechanisms. Exception might be documents of external actors such as audit institutions or the media. However, audit is in the Ministry of the Flemish community in its infancy. The problem with the media is its selective attention. They usually focus on some policy sectors more than others and they overweigh the negative over the positive. Therefore, we opted for face-to-face interviews with the managers of the sections.

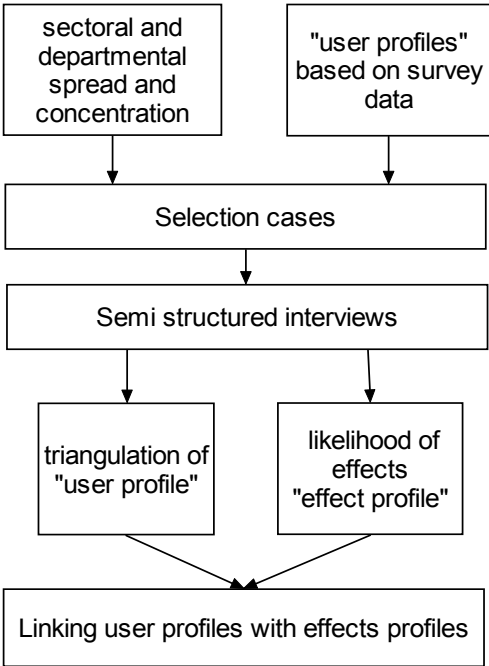


Figure 39: Overview of the methodology and data

The twelve sections of the Ministry of the Flemish Community were selected based on two criteria. First, we required combination of concentration and spread over policy sectors and departments. This should allow for comparison within and between sectors and departments. This also allows for us to

keep constant the characteristics of the department of which the sections are a part. Yet, we do not lose the possibility to compare across departments<sup>82</sup>.

Secondly, we selected sections that have moderate to high experience with performance measurement. Obviously, this is a necessity for studying the effects of performance measurement. We determined the experience with performance measurement by using the survey of 138 sections of the Ministry of the Flemish Community (see p. 121). In order to differentiate between the uses of performance measurement, we selected sections with different user profiles. The user profiles are based on the categories of use.

In order to establish the initial user profiles, we used survey data. Based on indications of use, scores on three categories of use were calculated. The indications are represented in Table 37. The calculation is straight. Every indication gets a standardized score, which is 1 or 0 (present not present), or the score on a scale divided by the maximum value of the scale. Next, we add up the scores and divide the sum by the total indications for the category (4). In case of ordinal scales, missing values were replaced by the series mean. In case of nominal (1/0) missing values are set to be 0. We assume that in general if organizations do have for instance self-assessment or indicators in an annual report, they will report on that. If they do not have it, they may not answer the question for reasons of social desirability.

---

<sup>82</sup> The organic structure of the Ministry of the Flemish Community is represented on page 121. The Ministry consists of seven departments. Each department has a number of administrations and sections. The sections are the most elementary decision making unit.

<i>Research function</i>	<i>Internal Management</i>	<i>Accountability</i>
Is the section involved in policy preparation or evaluation?	To what extent does manager of the section use performance information of the section in the annual evaluation of the professional staff?	Is performance information of the section used in the accompanying documents to the annual budget?
<i>Rescaling: 1 if there is involvement in preparation and/or evaluation. 0 if there is only involvement in policy execution.</i>	<i>Rescaling: the scale from 0 to 4 is divided by 4.</i>	<i>Rescaling: 1 for yes, 0 for no</i>
Is performance information of the section used in the minister's policy plans?	To what extent does the manager use performance information of the section in the annual evaluation of the other personnel of the section?	Is performance information of the section used in annual report of the section?
<i>Rescaling: 1 for yes, 0 for no</i>	<i>Rescaling: the scale from 0 to 4 is divided by 4.</i>	<i>Rescaling: 1 for yes, 0 for no or no annual report</i>
Is the performance information used in Self Assessments?	Is performance information of the section used for allocation of resources?	Is performance information of the section used in monitoring policy contracts with agencies?
<i>Rescaling: 1 for yes, 0 for no</i>	<i>Rescaling: the scores on the scale from 0 to 5 are divided by 5.</i>	<i>Rescaling: 1 for yes, 0 for no</i>
Is performance information used for adjusting policies?	Is performance information of the section used for the steering of teams?	To what extent is performance information of the section used in the annual evaluation of the manager of the section?
<i>Rescaling: the scores on the scale from 0 to 5 are divided by 5.</i>	<i>Rescaling: the scores on the scale from 0 to 5 are divided by 5.</i>	<i>Rescaling: the scores on the scale from 0 to 4 are divided by 4.</i>
$\Sigma$ (scores) /4	$\Sigma$ (scores) /4	$\Sigma$ (scores) /4

Table 37: Indications from survey data for the initial user profiles

The combination of the two selection criteria (sectoral and departmental spread and concentration, and the initial user profiles) led to the following selection.

<i>Policy Sector</i>	<i>Department</i>	<i>Section</i>	<i>Research</i>	<i>Management</i>	<i>Accountability</i>	<i>Profile (++) ≥ 0.75; 25 -75; 0 &lt; 25)</i>
Education	Department of education	EDU_1	0,75	0,49	0,44	<b>Research</b> (++) , Management (+) Accountability (+)
		EDU_2	0,05	0,75	0	Research (0), <b>Management</b> (++) Accountability (0)
		EDU_3	0,65	0,2	0,25	Research (+), Management (0) Accountability (+)
		EDU_4	0,92	0,39	0,75	<b>Research</b> (++) , Management (+) <b>Accountability</b> (++)
		EDU_5	0,65	0,1	0,5	Research (+), Management (0) Accountability (+)
Spatial planning and housing	Department of Environment and Infrastructure	SPH_1	0,42	0,28	0,25	Research (+), Management (+) Accountability (+)
		SPH_2	0,4	0,28	0,31	Research (+), Management (+) Accountability (+)
Waterways and marine affairs	Department of Environment and Infrastructure	WATER_1	0,25	0,25	0	Research (+), Management (+) Accountability (0)
		WATER_2	0,7	1	0,63	Research (+), <b>Management</b> (++) Accountability (+)
Culture	Department of Welfare, Public Health and Culture	CUL	0,92	0,36	0,75	<b>Research</b> (++) , Management (+) <b>Accountability</b> (++)
Public Health	Department of Welfare, Public Health and Culture	HEALTH	0,7	0,83	0,69	Research (+), <b>Management</b> (++) Accountability (+)
Media	Department of Science, Innovation and Media	MEDIA	0,55	0,2	0,5	Research (+), Management (0) Accountability (+)

Table 38: Selection of the cases based on policy sector, parent department and initial user profile



The survey data on the one hand and the semi structured interviews on the other should allow for both a triangulated user profile and an assessment of the effects of performance information. The triangulation of the user profile will be based on a set of items that refers to the uses of performance information formulated by Behn (2003) (Table 5). The effects are assessed from a list of effects are described on page 195. Note that we only studied the effects that have an impact on the output. The interviews were not taped. Yet, a report of each interview was drafted and sent to the managers for feedback in order to increase construct validity (Yin 1994). Most of the managers suggested improvements to the reports. The analysis is based on the final reports. Finally, the link between the profiles should provide insight about the central hypothesis. Does use determine effect?

<i>Eight purposes that public managers have for measuring performance</i>		
<b>Purpose</b>	<b>For which questions can performance information be useful?</b>	<b>Category</b>
Evaluate	How well is my organization performing?	Research
Control	How can I ensure that my subordinates are doing the right things?	Management
Budget	On what programs, people, or projects should my organization spend money?	Management
Motivate	How can I motivate people to align oneself with the objects of the section?	Management
Promote	How can I convince superiors, politicians, interest groups and the media that my organization is doing a good job?	Accountability
Celebrate	How to show good performance of the section to the employees?	Accountability
Learn	Why is what working or not working?	Research
Improve	What exactly should we do differently to improve performance?	Research

Table 39: Eight purposes that public managers have for measuring performance (Behn 2003) - items in the semi-structured interviews

Finally, we stress that this study is qualitative research. Qualitative research goes beyond how much there is of something to tell us about its qualities (Miles and Huberman 1994). Apart from the case selection, the study is based on approx. 10 page interview reports. This is rich data. Yet, in order to make sense of the data, we will be counting. This is not in contradiction with the qualitative nature of the study. Even in qualitative research, a lot of counting goes on in the background. Miles and Huberman (1994: p.253) give three reasons for using numbers in qualitative research. First, numbers allow to better see what data you have. Numbers are more economical. Second, counting may be done to verify a hypothesis. It allows identifying the cases that confirm or reject the hypothesis. Thirdly,

they assert that counting protects from bias. It is more difficult to mould research findings - intentionally or unintentionally<sup>83</sup>.

## 9.6. Results

### 9.6.1. triangulation of the user profiles

The profiles that we derived from the survey and that were used in order to select the cases will now be examined. Triangulation in social research is the combination of different methods, methodological perspectives or theoretical viewpoints (Miller and Brewer 2003). The metaphor of the stability of a tripod is used to assert that the result of using varied approaches is a net gain. Four methods of triangulation can be identified (Denzin 1978). First, *between-method triangulation* uses different research techniques - usually a qualitative and a quantitative one. Second, *within method triangulation* uses different variations within a technique (for instance different scales in a survey) to measure the same thing. Thirdly, *investigator triangulation* implies that two or more researchers independently study the same phenomenon. Fourthly, *data triangulation* is about using different data blocks but the same methodology. We are applying between method triangulation. The main idea is to compensate the strengths and the weaknesses of both quantitative and qualitative methods. The triangulation is based on the semi-structured face-to-face interviews.

Table 40 represents the triangulation schematically. The first column is an acronym of the sections. The code refers to the policy sector: education (EDU), spatial planning and housing (SPH), waterways and marine affairs (WATER), culture (CUL), health (HEALTH) and media (MEDIA). The second column repeats the initial profile that was used for selecting the cases. Each case may have a high, moderate or low profile on the three categories of use - research, management and accountability. The third column contains the updated profile based on the interviews. The fourth column provides the degree of convergence between the initial and updated profile. The measure gives the number of actual corrections over the number of potential corrections. For instance a change from + to ++ is counted as one correction. A change from ++ to 0 is counted as two corrections. The latter correction is more drastic. EDU\_1 potentially has four corrections. RES (++) can have a maximum of two corrections to RES (0). MGMT (+) can have maximum one correction to MGMT (0) or MGMT (++). The same goes for ACC. The score 1/4 means that the initial profile is 25% different from the updated profile. Finally, we indicated the direction of the correction. A positive correction indicates that the

---

<sup>83</sup> Miles and Huberman (1994: p253) state that 'people - including researchers - tend to overweight facts they believe in or depend on, to ignore or forget data not going in the direction of their reasoning, and to "see" confirming instances far more easily than disconfirming instances.'

initial profile was updated towards more measurement. A negative correction means that the initial profile was updated towards less measurement.

<i>Section</i>	<i>Initial Profile</i>	<i>Updated profile based on face to face interviews</i>	<i>Degree of Divergence</i>	<i>Direction</i>
EDU_1	RES (++) , MGMT (+) ACC (+)	RES (++) , MGMT (++) , ACC (+)	1/4 0.25	+
EDU_2	RES (0) , MGMT (++) ACC (0)	RES (+) MGMT (++) ACC (0)	1/6 0.33	+
EDU_3	RES (+) , MGMT (0) ACC (+)	RES (0) MGMT (++) ACC(++)	4/4 1	++++
EDU_4	RES (++) , MGMT (+) ACC (++)	RES (+) , MGMT (++) ACC (++)	2/5 0.4	+ -
EDU_5	RES (+) , MGMT (0) ACC (+)	RES (+) , MGMT (0) , ACC (+)	0/4 0	
SPH_1	RES (+) , MGMT (+) ACC (+)	RES (+) , MGMT (0) , ACC (+)	1/3 0.33	-
SPH_2	RES (+) , MGMT (+) ACC (+)	RES (+) , MGMT (0) , ACC (0)	2/3 0.66	--
WATER_1	RES (+) , MGMT (+) ACC (0)	RES (+) , MGMT (0) , ACC (+)	2/3 0.66	+ -
WATER_2	RES (+) , MGMT (++) ACC (+)	RES (+) , MGMT (++) , ACC (+)	0/4 0	
CUL	RES (++) , MGMT (+) ACC (++)	RES (0) , MGMT (0) , ACC (+)	4/5 0.8	----
HEALTH	RES (+) , MGMT (++) ACC (+)	RES (0) , MGMT (0) , ACC (+)	3/4 0.75	---
MEDIA	RES (+) , MGMT (0) ACC (+)	RES (+) , MGMT (+) , ACC (+)	1/3 0.33	+
			AVG = 0.46	

Table 40: Triangulation of the survey based user profiles by means of semi-structured interviews

EDU\_1 has a strong management profile, which remained underexposed in the survey profile. The performance information is widely available within the organization, is used to measure workload and to reallocate resources between teams. In addition, the performance information plays a role in rewarding financial bonuses. Qualitative judgments however are more important for the latter purpose. The research profile is also strong. Operational learning is definitely present. The development of the measurement system coincided with a reorganization of the organization. It is acknowledged that the management information potentially has high policy relevance. Steps (such as extra registrations) are taken to further develop policy relevance.

The management profile of EDU\_2 is confirmed. Allocation of personnel, comparing performance of people working at home versus people working at the workplace, plans for experiments with remunerations of teams, follow up of the progress of files endorse the prominence of the management profile. The accountability profile is low. It is stated that political interest in the information and the activities of the section is low and that the main task of the manager is to keep the minister out of the media. Quantitative information is seen as less useful for accountability purposes. This would require substantial explanation of the performance information for which there is seldom an opportunity. The

research role is moderate instead of low. There is evidence that the information is used to evaluate some policies and operations.

The initial profile of EDU\_3 appeared seriously flawed. The face-to-face interview gave a different picture of the organization. Whereas the survey pointed to a moderate use for research, the interview provided evidence of a high accountability and management use. The performance data is used for internal control and allocation of resources. In addition, it is experienced that those sections that measure performance come under attack more quickly than sections that do not measure performance. Moreover, performance information is not found helpful for supporting budget requests. On the contrary, a qualitative description with some *flou artistique* is said to be more fruitful.

EDU\_4 is using performance information intensely. Unlike the initial profile, the utilization focuses on internal management. The outliers are looked at and discussed with teams and/or individual employees. Requests for additional resources by team leaders need to be substantiated with performance information. Use for the research function is moderate. Although it is acknowledged that performance information may be useful to evaluate the performance of the section, the main organizational learning effects are expected from qualitative self-assessments. Policy learning is seen as a responsibility of politicians. The accountability function is high, although not seen as threatening.

The initial profile of EDU\_5 is accurate. There is modest use for research and learning. The performance information feeds mainly into the policy process. There is a substantial interest from the political level (the cabinet of the minister of education and parliamentary questions). Organizational processes are only limitedly reviewed based on performance information. There is low use for management purposes. Some control, but without sanctions can be based on the rank-and-file employees. Accountability purposes are moderate. Chiefly, higher executives require performance information.

SPH\_1 has moderate use of performance information. The initial profile appears accurate. First, research and learning takes place mainly by looking in depth at outliers. In addition, regional offices are compared. The learning is more at an operational output level. Although policy options resonate in the results, it is seldom the case that the data is used for drawing policy lessons. Secondly, accountability purposes are also moderately present. In particular, the data is used to demonstrate understaffing. However, the results of the effort are seen as insufficient. The activities of the section are not a political priority. The management profile may be overestimated in the initial profile. There is only evidence of some work planning based on performance information of third parties, i.e. municipalities.

SPH\_2 is a low user of performance information. The section does only have limited information on its own operations. The initial profile overestimated this. Management and accountability use is lower. The use for research and learning is moderate. It involves predominantly information for policy purposes. In fact, data gathering is an important activity for the section.

WATER\_1 resembles SPH\_1. It is also a section that collects a lot of information and has a policy orientation. However, performance information is more used for accountability. The indicators are a part of a Balanced Scorecard, which is a departmental initiative. The indicators are discussed in a meeting with section heads on a three monthly basis.

The initial profile of WATER\_2 is accurate. This section uses performance information intensely, and mainly for management purposes. Time registrations and the number of projects in time point to the use of information for internal control. Measurement is also needed to calculate the cost of projects. Accountability and research and learning purposes are moderate. The annual report is an important tool for accountability. Research and learning based on the information is present for both policy and management purposes.

CUL's initial profile needs significant modification. The use of quantitative information is relatively low. The section has a moderate accountability profile. This is done through the annual report, which is widely spread amongst the stakeholders. Yet, it is said that for convincing stakeholders personal contacts are probably more important. The management function relies more on qualitative information, rather than quantitative. Financial bonuses for instance are not rewarded based on measurement, but on a qualitative assessment of extra efforts and responsibilities. The research function is also low. Although measurement information may be used to justify decisions ex post, it is seldom the case the performance information determines a decision ex ante.

The initial profile of HEALTH also needs modification. Belief in the utility of performance information for research and learning, and for management control is very low. There is more reliance on trust, peer control, and qualitative assessments. It is doubted whether the costs of more measurement outweigh the benefits. Also, it is assessed that the potential for policy learning does not materialize because of the lack of capacity to evaluate. Measurement efforts are for an important part an external obligation. In particular, the role of higher hierarchical levels in developing more measurement stands out. Therefore, the score on accountability is accurate.

Finally, the initial profile of MEDIA is reasonably accurate. The section is a modest user of performance information for a wide variety of uses. Timeliness in coping with dossiers is an important indicator for higher hierarchical levels. This is an indication of the accountability function. Policy relevant data is collected for research and learning purposes. Management control of dossiers is partly

based on performance information. Two weekly listings give an indication of the backlog. Yet, the relatively small size of the section does allow for personal contact and more informal control.

The overall degree of divergence between the initial and the updated profile is 0.46. This means that the initial profile, based on three times four indications of respectively use for research and learning, management and accountability is approximately 50% correct. Five profiles out of twelve were significantly corrected. Twelve times, the direction was towards less measurement. Nine times, it was a positive correction - i.e. more measurement.

9.6.2. Which effects are showing?

We put the question as follows: “which effects of performance measurement are the most realistic?” We did not ask directly whether they did occur or not. The respondents being managers of a section, this would be increasing the propensity of socially desirable answers. The assessment of the degree of realism allows for an assessment of previous experiences as an evaluation of current practices and a forecast on future expectations. The description of the effects is represented in Table 41.

Effect	Interview question: which effects of performance measurement are the most realistic?
Hypertrophy	The measured services or products are inflated in order to obtain good scores.
Atrophy	Quantity goes at the expense of quality
Tunnel Vision	The organization loses sight of the activities that are not measured
Skimming (cherry picking)	Only or mainly the easy cases are dealt with, in order to have quick results.
Gaming	Output is deliberately lowered in order to avoid raising standards
Ossification	The focus on quantitative information immobilizes the organization. It hinders experimentation.
Myopia	There is a focus on the short term at the expense of the long term
Sub-optimization	Indicators push for an optimal organizational output which is not necessarily an optimal societal output
Complacency	Soft targets make organizations lean back. They are not sufficiently ambitious.
Total loss effect	The “desperate cases” are no longer given priority (they are parked) in order not to delay other cases.

Table 41: The phrasing of the research question and description of the effects

Table 42 represents the findings. We made an assessment based on the reports of the interviews using an ordinal scale, which for each effect ranges from 0, not realistic, over 1, somewhat realistic, 2 realistic and 3 very realistic. The vertical sum gives an overall assessment for each section. Horizontally, four measures are considered. First, we give the average score for each effect. Secondly, the distribution over the effects is calculated. The sum of the averages is equaled to 100%. Next, the average for each effect is divided by the total average. This gives an indication of the importance of an effect vis-à-vis the other effects. When negative effects occur, one chance out five it is hypertrophy - the inflation of measured activities. The probability that it is the “total loss” effect is 3%. Obviously, the

calculation of the probability is based on the twelve interviews with managers of sections in the Ministry of the Flemish Community. The results cannot be generalized beyond this context without new interpretation and empirical data. The third assessment gives the average in relation to the maximum score per effect. The maximum score for each effect is 3. The quotient of the average and the maximum gives an indication on how much of the potential impact actually risks materializing. It may be understood as the sense of reality of the effects. Finally, the last column represents the number of cases where the effect is not zero. This gives an idea of the spread of the effects over the sections.

Effect	Section												AVG	over effects	per effect	Count (N >0)
	a	b	c	d	e	f	g	h	i	j	k	l				
1. Hypertrophy	2	0	2 <sup>84</sup>	1	0	2 <sup>85</sup>	2	1	3 <sup>86</sup>	2	1	3 <sup>87</sup>	1,58	20%	53%	10
2. Atrophy	2	2	3	1	2	0	2	0	2	0	2	1	1,42	18%	47%	9
3. Tunnel Vision	2	0	1	1	2	0	0	0	2	1	0	2	0,92	11%	31%	7
4. Skimming	0	0	2	1	0	0	1	0	2	0	2	2	0,83	10%	28%	6
5. Gaming	2	0	1	1	2	0	0	0	1	1	0	1	0,75	9%	25%	7
6. Ossification	2	0	0	0	2	0	2	0	0	0	2	1	0,75	9%	25%	5
7. Myopia	0	2	0	1	2	0	0	0	0	0	2	0	0,58	7%	19%	4
8. Sub-optimization <sup>88</sup>	0	0	2	1	0	0	0	0	1	0	0	2	0,50	6%	17%	4
9. Complacency	0	0	0	0	2	0	0	0	2	1	0	1	0,50	6%	17%	4
10. Total loss	1	0	0	1	0	0	0	0	0	0	0	1	0,25	3%	8%	3
<i>Sum</i>	11	4	11	8	12	2	7	1	13	5	9	14		100%		

0 not realistic, 1 somewhat realistic, 2 realistic, 3 very realistic

Table 42: The effects of performance measurement: assessment per section, average, count, distribution over effects, and probability ratio per effect

The most likely effects are hypertrophy and atrophy. The probability that the measured output is pushed up is seen as high. It should be noted that several managers explicitly mentioned that this is not necessarily a bad thing. As a complement, it is feared that quality of output may suffer from this focus on quantitative measurement. Tunnel vision is the third likely effect. Here, the quality of output does not erode. Yet, activities that are not measured are lost out of sight. Skimming and gaming are in the middle of the range. These responses are about strategically managing respectively the intake and

<sup>84</sup> It is acknowledged, realistic effect. However, it is not seen as negative one. There is what is called 'a platoon (peleton) spirit'. Nobody likes to be an outlier, in particular on the negative side.

<sup>85</sup> See note 71; it is said that "more is sometimes better, as long as it concerns essential things".

<sup>86</sup> See note 71 and 72; there are two consequences of measurement: output is 'discovered', (previous activities are now recognized as output, whereas before they passed unnoticed) and real output is higher as well

<sup>87</sup> see note 71: more output is often intended

<sup>88</sup> Several respondents answered that the optimal societal output is a political responsibility, and therefore it is hard to assess whether measurement may to a suboptimal output. Since these respondents do not see

the output level. Ossification, the loss of the ability to experiment, is also seen as moderately likely. The less likely effects are sub-optimization, complacency and myopia. These effects relate to defective target setting, i.e. socially undesirable targets, insufficiently ambitious targets, and short term targets without long term view. The likelihood of the effect whereby the hopeless cases are set aside (total loss) is virtually absent.

### 9.6.3. Linking user profiles and effects

Figure 40 links the assessment of use (Table 40) with the assessment of the effects (Table 42). The vertical axis gives for each section the sum of the scores for the 10 effects of performance measurement (which ranged from 0, not realistic to 3, very realistic). The horizontal axis gives the sum of the three assessments of use, i.e. use for research and learning, use for internal management and use for accountability. The total is broken out for the three uses (R; the first segment, M; the second segment and A: the third segment). The maximum score is represented by the upper line. Obviously, this is a virtual case.

---

societal optimization as their responsibility, it is assumed that the effect is not realistic.



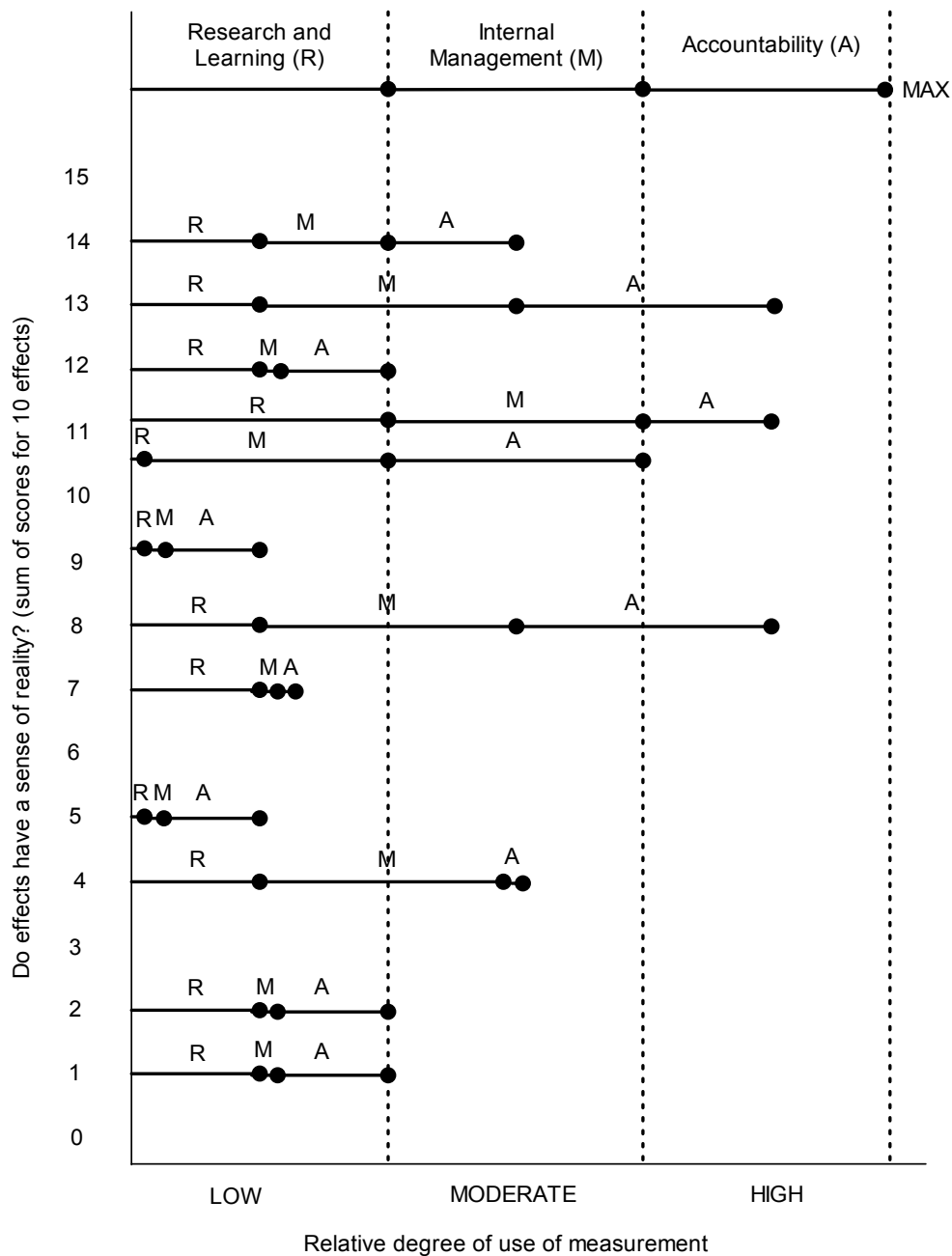


Figure 40: Linking use and effects

At this point, we can return the hypotheses.

*H1. High use will lead to high effects: organizations that are measuring more will attach a higher sense of reality to the effects.*

Yes, there is a tendency for high users to see the effects as being more likely. The four sections that have the highest user profile are in the upper half of the scale. They score respectively 8, 10, 11, and 13. Four out of six low users are in the lower half of the scale (1,2,5,7). Yet, two out of six of the low

users do feel that many effects have a high sense of reality (9 and 12). Moreover, the moderate users are scattered on the scale (4,11,14).

Let us look at the cases that do not confirm the hypothesis (the two low users with high effects and the moderate user with high effects). In one case, the manager was skeptical about the added value of measurement. Nonetheless, the pressure of higher echelons in the Ministry of the Flemish Community to implement measurement systems was clearly perceived. This may have evoked a reaction of defense against performance measurement. In main factor behind the two other cases in all probability is the example of other sections. Both sections have intense dealings with organizations that are high users. They feel the effects are realistic because they see things happening around them.

*H2 high use for accountability will lead to the highest effects, no matter what the utilization for research or management is.*

The three sections with the most outspoken accountability profile have relatively high scores on the effect side (13, 11, 8). This confirms the hypothesis. Managers of sections with a strong use for accountability attach a greater sense of reality to the effects. Yet, the sections with a moderate accountability profile are scattered along the scale. It should be noted that accountability based on performance information in Flanders is not as heavily and externally enforced compared to the Anglo-Saxon world<sup>89</sup>. Therefore, a high accountability profile in the Flemish context would probably be a moderate one in other contexts such as the United Kingdom.

*H3 high use on management, but low on accountability will lead to moderate effects, no matter what the utilization for research is.*

Two sections have strong management profile, but a moderate accountability profile. One has a low score on the probability of effects (4); the other has a high score (11). The other three sections with a high management profile have a strong accountability profile too. Interestingly, the former two sections are in the same policy sector, have comparable activities and size. The main explanation we saw for the different scores was the personality of the manager. Not every manager is as inclined to identify effects.

---

<sup>89</sup> The United Kingdom has a strong accountability culture for many aspects of government: see for instance Andrews (2004) and Game (2005) for an analysis of local government; Wilson (2004) and Wiggins and Tymms (2002) for schools; Marshall, Shekelle, Leatherman and Brook (2000) and Marshall Shekelle Davies and Smith (2003) for the health sector in both UK and USA.

It should also be noted that those sections that have a low management profile tend to have low scores for the effects. Exceptions are the section (high on management, low on effects) we mentioned in the previous paragraph and the two sections that are confronted with hierarchical pressure and experiences of peer organizations. The effects of the management profile seem to parallel the effects of the accountability profile. The pressure for goal displacement seems to arise as much from management applications as from accountability purposes.

*H4 high use for research will lead to low effects, under the condition that use for management and accountability is low.*

In general, the research and learning profile is not very outspoken. Only one section has a strong research and learning profile. Attempts at learning are being made for both policies and operational matters. The other sections only moderately use performance information for research and learning.

Four sections combine a moderate research and learning profile with a low management profile and a moderate to low accountability profile. Yet, they are evenly distributed over the effect scale. Two sections have very low scores on the effect scale (1,2). One has an average score (7) and one has a high score (12). The latter organization however has peer organizations with a high management and accountability profile.

Several sections combine a research and learning profile with a moderate to high management and accountability profiles. The single section with a strong research and learning profile has a strong profile on management and a moderate accountability profile. Two sections with a high accountability profile have a moderate research profile too. The thesis that high-impact use drives out low-impact use cannot be confirmed based on this data.

Again, it is important to notice that this result applies on the Ministry of the Flemish Community in the early 2000s. At the moment of the research, external actors or requirements do not enforce measurement. The only pressure arises from higher echelons in the Ministry of the Flemish Community. The departments are just commencing the implementation of management scorecards (usually a Balanced Scorecard approach). The sections typically have to provide some Key Indicators for these scorecards. Hypothetically, another picture will show when the superiors of the section managers will start to use the management scorecards for steering and controlling the sections. A follow-up research after some years may give an insight about the evolution.

Thus, we have to reconsider the hypothesized function (see Figure 38: Central hypothesis of measurement effects: breakdown for categories of use on page 202). Figure 41 provides the revised picture. The sections with a research profiles are equally spread across the effect scale. The effects of

research and learning are fixed ( $\Delta = 0$ ). They do not seem to increase together with increased use<sup>90</sup>. For accountability and research profiles, the effects seem stronger with higher use. Yet, unlike the initial hypothesis, the accountability profile does not have a more profound impact on the sense of reality that managers attach to effects compared to the management profile ( $\Delta \text{ACC} = \Delta \text{MGMT}$ ).

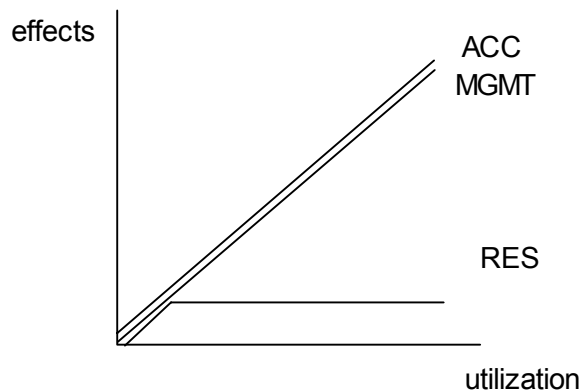


Figure 41: Effects and uses in the case of the Ministry of the Flemish Community

## 9.7. Opportunity and motive

The results demonstrate that the category of use has an impact on the effects. Yet, use does not seem to be the only factor that explains the effects of performance information. The explanations that managers provided for their assessment of the likelihood of the effects may give some insight into these other factors. The respondents argued why they see an effect as more or less likely and how they remedy the effects. A useful distinction for explaining the effects of performance measurement on organizational behavior is between the motive and the opportunity for goal displacement. Is it useful and is it possible to change output because of performance measurement?

The categories of use are providing a motive for changing behavior. Up to this point, the line of reasoning is that accountability and management purposes have a more pressing and direct impact on the organization. The direct stakes in terms of budget allocations, bonuses or sanctions are higher. Therefore, this may provide a motive for altering behavior in order to obtain good scores on performance indicators.

---

<sup>90</sup>We depicted the function for the research and learning profile also as starting in the origin. Not using performance information will not cause effects of performance information. Yet, our observations suggest that the effects of the research function flatten quickly.

Yet, a motive is not enough. There also needs to be an opportunity for altering output. In many cases, organizations do not have this possibility. Table 43 gives the explanations that managers provide for the occurrence of effects. They mentioned some external limiting factors as well as some strategies for counteracting them.

Table 43: The motive and the opportunity for goal displacement due to performance measurement: limiting factors of the effects

<i>A. Opportunity – limiting factors</i>	<i>B. Motive - limiting factors</i>
<ul style="list-style-type: none"> <li>✓ No control over intake</li> <li>✓ Characteristics of cases</li> <li>✓ Networks within the administrative sphere</li> <li>✓ Networks within the policy sector between administration, politics and societal actors.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Habit formation and increased predictability</li> <li>✓ Consequential loop (boomerang effect)</li> <li>✓ Intrinsic motivation</li> </ul>
<i>C. Opportunity – counteracting strategies</i>	<i>D. Motive – counteracting strategies</i>
<ul style="list-style-type: none"> <li>✓ Information and Communication Technology (ICT): control and standardization</li> <li>✓ Double checking (of a sample)</li> <li>✓ Measuring quality (e.g. through client surveys)</li> <li>✓ Qualitative assessment of quality</li> </ul>	<ul style="list-style-type: none"> <li>✓ 100% norm</li> <li>✓ Flexible dealing with targets</li> <li>✓ No targets</li> <li>✓ Phrasing of the targets in general terms</li> </ul>

and counteracting strategies

A. Four factors, which limit the possibility of altering output, were identified. First, several managers mentioned that they did not have an impact on the intake. Therefore, effects such as cherry picking, hypertrophy, and gaming are not possible, even if they would want to do this. Secondly, the characteristics of the case may limit the opportunity. This is for instance the case for the total loss effect. Putting a case aside is only possible when the importance for individual beneficiaries is limited. Therefore, it may be easier ‘to park’ a case if several people are minimally affected then when few people are affected heavily. Delaying a delayed train even more is easier than setting aside a dossier for subsidizing a school or paying a wage.

The third and the fourth limitations of the opportunity are caused by peer insight into the production process of the information. Thirdly, the opportunity may be limited because of informal networks inside the administrative sphere. One manager mentioned that even in a context of performance contracts, the effects measuring performance are not likely because “one knows one another too well”. Fourthly, characteristics of the policy sector may also affect the opportunity for instigating effects. In general, the section is one actor in a triangular relationship between administration, politics and society. The latter are the organizations, companies and citizens who are affected by or play a role in policies. In some policy sectors, the distance between societal actors, politicians and the administration is closer relative to others. This is particularly the case when the sector is well organized, with strong umbrella organizations. In this case too, the opportunity to alter output to obtain favorable scores is reduced because of peer insight into the production function of the organization.

B. Not only the opportunity is limited. The managers also provided insight about why there is no interest for their sections to bring about these effects. First, a manager of a section that intensely uses performance information mentioned that the effects decrease over time. The increasing experience in working with performance information leads to habit formation. The use of performance information becomes more predictable and - in this case - it also becomes less threatening. It should be noted that another section with high use profile reports exactly the opposite. It is expected that when the system becomes more predictable, more effects will occur. The system is predicted to erode as an instrument for internal management, because it will lose its ability to diversify. In the literature, this phenomenon is known as the performance paradox<sup>91</sup> (Meyer and Gupta 1994; Van Thiel and Leeuw 2002). The experiences of the two sections show that habit formation and increased predictability may work in different ways.

Secondly, the motive may be affected because of a consequential loop. The section, or even the individual employee, that alters its output may be directly affected in a later stage. For instance, pursuing quantity at the expense of quality may result in a higher workload because more appeals are lodged with administrative courts. As one manager puts it: "low quality always gets back on our desk". Inferior quality gets back like a boomerang. Obviously, this is not always the case. A noteworthy counterexample is low quality in paying allowances. When low quality means that decisions are made in favor of the beneficiary, cases will not return. If judgments tend to favor the recipient at the expense of the state, then low quality may lead to even more quantity.

Thirdly, several sections mentioned that the intrinsic motivation is dominant. Therefore, there is no need for skewing performance indicators. This was explicitly mentioned in highly professionalized services where most of the staff have a high degree of schooling. The esprit de corps is stronger than the pressure to obtain high performance scores. Dimaggio and Powell would call this normative isomorphism (1983). This is the adaptation of organizational behavior to norms. These norms are mainly determined through professionalization. There are two sources of professionalization. On the one hand, there is the formal education and legitimization of cognitive base. On the other hand, the growth of professional networks and associations also exerts normative isomorphic pressure. In this case, the process of normative isomorphism countervails the pressure of performance information.

---

<sup>91</sup> The performance paradox states performance indicators have the tendency to run down over time (Van Thiel and Leeuw 2002). They lose their value as measurements of performance because they can no longer discriminate between good and bad performers. Meyer and Gupta (1994) attribute this effect to four processes; positive learning, perverse learning, selection and suppression. Gormley and Weimer (1999) provide an overview of scorecards in the USA in diverse policy contexts such as schools, colleges, hospitals, day care centres, nursing homes, local governments and state governments. Coe (2003) reports on US report cards about the environment, K-12 education, higher education, fire protection, health and social welfare, economic development, and state management and policy.

The section managers also provided some insight into how they dealt with these effects. The counteracting strategies are focusing on the opportunity to alter output or the motive for altering output.

C. First, sections relied on information and communication technology (ICT) to reduce the opportunity for skewing performance measures in two ways. In the first way, ICT was used to control processes. Computerization allows for the collection of data that is broken down to the individual employee and the individual file. Individual performance can be assessed. It should be noted that there is some reluctance in using performance measurement information in this way. Precisely because of the potential effects on the measurement system itself, several managers explicitly communicated to their staff that they would not use performance information for control of individuals. Those who did use it for the purpose of individual control stressed that they only looked at outliers and provided enough room for explanation and interpretation. Secondly, ICT is used for aligning interpretations. This standardization limits the opportunity for misclassification - deliberate or not. Secondly, sections use traditional control techniques. A popular technique is double-checking a sample of files. If the probability of being controlled is perceived to be significant, then this will reduce the opportunity of skewing output. Thirdly, several managers mentioned an explicit attention to quality. Two strategies are followed. Some sections decide to measure quality. Measurement defects then are counteracted by more measurement. Client surveys are an example. Other sections resort to qualitative assessments of quality. This is can be done by having conversations with staff and stakeholders.

D. The strategies for taking away the motive all are directed towards target setting. The most obvious strategy is not to set a target, and just let the figures speak for themselves. Yet, without a target, people will rely on other ways of sense-making when confronted with the data. The most likely point of reference will then be last year's performance or performance of comparable units. If this is the case, the motive gets back in by the backdoor, but remains implicit. Another strategy is to formulate an absolute standard (100%). For instance, the norm in one section was not to make any mistakes on commands to pay wages. The impact of paying a wage on the receiver probably justifies this stance. However, as a yardstick, the target is not useful. Again, the implicit standard is not clear. Still another strategy is to phrase the targets in general terms. Rather than a precise figure, the targets are formulated in terms of an increase or decrease of the results. Still, there is a target but not as strict as a numeric standard. Moreover, the standard is made explicit. In general, almost all managers stress that they deal flexibly with the targets. Not meeting a target is not a problem when good explanations are provided.

## 9.8. Conclusion

In this section, we had four objectives. First, we wanted to make an inventory of the main effects, which are found in the literature. In total we identified seventeen effects. These effects are of two kinds. Seven effects concern measurement errors. The inputs, activities and outputs of organizations remain the same. Yet, the representation in numeric terms by means of measurement does not correspond with reality. For example, when the counted number of clients is higher than the real number, this will at first not affect the operations of the organization. Ten effects refer to change in the output of the organization. In this case, inputs, activities and/or outputs are altered in order to comply with the indicators. We focused the remainder of the study on the latter effects of performance measurement.

Secondly, we looked for a common denominator for the effects. The concept of goal displacement is useful in this regard. Goal displacement occurs when an intermediate value becomes a terminal value. Applied on the effects of performance measurement, this means that a good score on the indicators becomes the ultimate value. In some cases, this is the intended effect. Indicators are often formulated precisely to direct the ultimate values of organizations. Yet, in many cases these are unintended effects. The fact that goal displacement may explain both intended and unintended effects is an important characteristic. It may bridge the gap between research findings of 'believers and disbelievers' in performance information.

Thirdly, we gave some empirical insight about the relative importance of the effects. This assessment is based on semi-structured face-to-face interviews of managers for twelve sections of the Ministry of the Flemish Community. The most likely effects have to do with increasing output of the measured activities (hypertrophy), often at the expense of attention for qualitative dimension (atrophy) or services which are not measured (tunnel vision). The effects that affect the production process (ossification), the intake (cherry picking) and the output levels (gaming) are moderately likely. The effects that result from bad target setting are the less probable ones. Myopia refers to short term targets without a long term view. In case of sub-optimization, targets are not societal optimal. Complacency results from targets that are not ambitious enough. Finally, giving difficult cases less priority is the least expected effect.

Fourthly, we linked the effect-profile of the twelve sections with the profiles of use. The hypothesis was that the potential for goal displacement would increase with an increased utilization. Moreover, these effects were expected to be stronger for accountability than for management use, and for management than for research and learning use. In general, more use leads to more effects. Yet this effect is equally attributed to the accountability and the management use. Accountability of sections in Ministry of the Flemish Community does not provoke more effects than the use for management purposes. An explanation for this relatively weak impact of accountability use compared to



management use may be the overall accountability culture. The accountability culture in Flanders is not as dominant compared to the Anglo-Saxon tradition. Probably, we would have found a stronger impact in context with a stronger accountability culture. Finally, we found that the increase in effects of performance information cannot be attributed to the use for research and learning. The latter use is distributed over the range of effects profiles.

One conclusion might be that use of performance information for research and learning is the most preferable for it has the least effects. Yet, it is the least developed one in the sections of Ministry of the Flemish Community. Other uses are more heavily institutionalized in tools and techniques such as Balanced Scorecards, unit cost calculation, annual reporting and bonus systems for personnel. Moreover, not only the applications, but also the results of use for research and learning are the least tangible. As a result, research and learning may also be the preferential use for symbolic reasons. There may be a thin line between use for real research and learning and the non-use of the performance information. Since accountability and management do not seem to be incompatible with use for research and learning, a multi track development is feasible.

The effects of performance measurement on the an organization will in turn have an impact on the use of measurement information. The manipulation of output to accommodate measurement results will affect the qualities of the indicators. The performance paradox - performance indicators have the tendency to run down because over time they can no longer discriminate between good and bad performers - is the most renowned example (Meyer and Gupta 1994; Van Thiel and Leeuw 2002).

In this chapter, we looked at the effects of performance measurement information. We tried to put the occurrence of the effects in an organizational context. This is necessarily a critical study for the practice of performance measurement. Yet, since every way of seeing is a way of not seeing (Poggie 1965), we would like to end by suggesting another research issue that is almost absent within performance measurement research: the effects of not having performance information.



## PART 3: CONCLUSIONS



First, we recapitulate the conclusions from the empirical research chapters as well as the conclusions from the historical chapter and the literature study. Next, we identify some possible next steps in the research of performance measurement in organization.

## 10. Conclusions

### 10.1. Conclusions from the historical study of measurement in government

Measurement initiatives have been taken at several points in time. We identified fourteen movements that promoted more measurement in government. Yet, most of these movements fell victim of over-commitment. After the initial hype, organizational memory loss appears to take place. Thus, an important question comes to the fore. Are we reinventing the wheel over and over again, or is there genuine change?

Our reading of history would suggest the latter. Yet, change is not the path of glory which is often portrayed. Measurement gradually conquered the public sector. First, from very rudimentary to more sophisticated techniques. This change mainly occurred in the 19<sup>th</sup> century. Most of the contemporary performance management techniques can be found in an embryonic form in the practices of the New York Bureau of Municipal Research. In the 20<sup>th</sup> century performance measurement became increasingly integrated into the government core. This process culminated in the New Public Management Movement.<sup>92</sup> Another relative newness is that performance measurement initiatives have become a deliberate export product. PPBS was the first initiative that was exported worldwide.

Another finding of the historical study of measurement in government is the joint development of measurement by public and private sector actors. Many recent performance management techniques, in particular at organizational level, did come from the private sector. This transfer found its origin in the 1980s when a crises of the (occidental) public sectors was combined with an image of purity and efficiency in the (Japanese) private sector. The image of government like the private sector, which was evident in the advent of the NPM, obscures the perspective. Many initiatives have clear public sector roots. PPBS and the New York Bureau of Municipal Research are two of the more renowned examples. The common opinion that performance measurement is a typical private sector phenomenon that NPM - whizzes transferred to the public sector needs to be adjusted.

---

<sup>92</sup> Note that the evolution of the metrics for performance measurement still progressed significantly. For instance, the development of Data Envelopment Analysis is a significant step forward and yields results that have a large

In addition, it may be argued that performance measurement in government is not merely an Anglo-Saxon tradition. Several innovations have European roots, in particular in the development of measurement in the 19<sup>th</sup> century. In addition, interest in policy indicators is more sustainable in Europe. The initiatives of the social indicator movement were more enduring in Europe. European systems appear to be less tempted to make *tabula rasa*. Most likely, the political system is a major explanatory factor. In a majority system (sometimes combined with a spoils system), political and administrative elites are thoroughly renewed after a political shift. In coalition systems, this process is more gradual.

Finally, we identified four transformations in performance measurement over time:

1. From ad hoc to systematic. Initially, peripheral actors measured government performance, mainly to influence decision makers. Nowadays, measurement is done on a more regular basis .
2. From generic to specialized. The increasing specialization of policy sectors led to a more specialized supply and demand for information within policy sectors. The gap between the increasingly specialized supply and the generalist demand of political decision-makers is an increasingly difficult issue in performance measurement.
3. From general to professionalized. A third and parallel trend is the increasing professionalization of measurement. This trend has two dimensions. On the supply side of information, professionalization implies that measurement has increasingly become a profession. On the demand side of information, there is a more professional way of dealing with information.
4. From anecdotic to institutionalized. Increasingly, measurement became embedded in the management and policy making systems of governments in addition to legislation.

## **10.2. Conclusions from the literature Study**

The literature study used a supply and demand approach to categorize the research questions in performance measurement articles of the last 20 years. The following conclusions were drawn from the analysis.

---

practice potential. Yet, this is mainly an academic endeavor, which still needs transferal to practice.

First, many studies aim at improving measurement. Yet, descriptions of optimal performance measurement systems usually do not relate prescriptions to a specific demand/purpose. The underlying logic is that there is 'one best way' to organize performance measurement. We argue in our study that there are at least several best ways to organize performance measurement and that the utilization of performance information should dictate the design.

Secondly, not many studies look at the (in)compatibility of different uses. Quite often, there seems to be a division between believers and non-believers. Believers find that performance information can be used as a panacea for all organizational challenges of both management and policy. Non-believers take the opposite stance and condemn the whole performance measurement venture. One of the main arguments of non-believers is the problems with performance related pay. Other criticisms include the lack of impact on resource allocation and the overall limitation of performance measurement in a complex, political environment.

Thirdly, the studies of the effects of performance measurement face the same challenge. Some studies critically describe the negative effects of performance measurement. Other studies only focus on the beneficial effects of performance measurement. Yet, studies seldom take into account that performance measurement is a multi-dimensional concept. Thus, particular effects may be more likely because of a specific use of performance measurement. Moreover, the effects, both positive and negative, are seldom empirically studied. Evidence is mostly anecdotic or case specific.

Fourthly, methodologically, most studies frequently apply small N methodologies and literature studies. Large N studies are less frequent. Paradoxically, quantification in government is seldom studied in a quantitative way. Performance measurement, management and policy is too often heavily entangled with the politico administrative system to fit in the superficiality of survey research. This observation has cautioned us to allow sufficient time for questionnaire construction. We did use large N research since this is the only research strategy in order to get an idea of the breadth of measurement in government.

Fifthly, we also found that the focus of performance measurement research broadens from administration to politics and citizens. It should be noted that this is a broader focus, and not a different focus. Performance measurement research in the 1980s mainly focused on the administrative system. Increasingly, the involvement of politicians and citizens in producing and consuming performance information is added.

### 10.3. Conclusions from the first empirical research chapter: What makes organizations measure?

First, a distinction was made between the adoption and the implementation of performance measurement. The adoption is about ‘having’ performance measurement system. Organizations can have many or few organizational processes that rely on indicators. The implementation is about ‘doing’ performance measurement. Organizations can use the information more or less intensely. Thus, adoption deals with the breadth of performance measurement in organizations while implementation refers to the depth of measurement. Adoption and implementation are not seen as subsequent phases in a process. An organization can have a high value for implementation and low on adoption. This would mean that the organization does not have many measurement systems, but uses them intensely. We found that adoption and implementation correlate. Yet, the scatter was dispersed. All four profiles are realistic.

		Adoption	
		Low	High
Implementation	Low	No PM	Outward PM
	High	Inward PM	Full PM

Table 44: Performance measurement (PM) profiles

Next, we tested six hypotheses on why organizations would measure performance. We briefly summarize the results.

- Measurability of output has a positive correlation with both adoption and implementation. Measurability is more important for adoption, and less for implementation.
- Political interest does not affect adoption or implementation of performance measurement. Yet, many organizations see it as a potential hindrance.
- Large organizations have a higher adoption and implementation. This observation identifies an important practical point in the development of a measurement policy: how to design low-keyed performance measurement systems.
- The reduction of discretion plays a role in the implementation of measurement systems, not for adoption. The study of outliers is the most frequent approach.
- The lack of resources does not affect adoption or implementation. Many sections experience a lack of resources. Some cope with it, while others do not.
- Coupling with goals is not neutral to adoption, but significant for implementation.



## **10.4. Conclusions from the second research chapter: Does administrative supply meet political demand?**

Politicians too are tempted by the power of expression of numbers. MP's often ask for indicators on policy issues. Indicators are included in 52% of the MPs' questions to the executive and in 48% of the executive's answers.

Although there is quantity, the quality of the answers is more uncertain. Only 44% of the answers (with indicators) provide exactly those indicators that have been asked for. The four explicitly mentioned problems were (1) the availability of the data, (2) the lack of timeliness, (3) coordination problems, and (4) consolidation problems. Political demand should be formulated in earlier stages.

There are significant differences between policy sectors. In quantitative terms, the policy sectors 'mobility and public works', 'employment', 'welfare and public health' and 'housing' have more indicators and the policy sectors 'internal affairs', 'education' and 'culture' have fewer indicators.

In addition, these policy sectors differ in focus of the measurement. For instance, 'Mobility and public works' is input oriented, 'housing' is output oriented and 'environment' is more effect-oriented. Finally, in most policy sectors demand is higher than supply. However, foreign policy and employment show a significant oversupply.

The reason for these differences may be diverse. Structural features such as the extent of decentralization and fragmentation in a policy sector may explain the supply and demand function. In addition, the measurability of the main policy issues and outputs may be a second reason. Finally, different cultural factors and practices in distinct policy sectors may relate to a dissimilar indicator orientation. These different explanations may interact and reinforce each other over time.

Finally, the relevance of the policy sector as a level of analysis stands out. The policy sector is an important unit of analysis, situated between the individual and the organization on the one hand, and the government-wide level on the other. Performance measurement typically has or an organizational focus including the implications government-wide initiatives for organizations or an intra sectoral focus (in particular education and health). This focus should be supplemented with a cross-sectoral angle.

## **10.5. Conclusions from the third research chapter: What are the system requirements for different uses of performance information?**

Degrees of freedom are restricted along with the underlying function of the use of performance information. Accountability purposes heavily restrict the operational freedom of organizations. Internal management purposes restrict the internal freedom of subordinates of the managers, but do not affect the managers themselves. The research function has learning and improving as a purpose. Therefore, it does not restrict degrees of freedom. Performance information in this case is an enabler.

The multitude of design parameters shows that a good fit between supply and demand of performance information is, only in part, a matter of the right quantity of performance information. The main challenge is to provide the right quantities with the right quality. The diverse uses of performance information require an adapted design of the measurement system. The adaptation involves all aspects of a measurement system, and not only the selection of the performance indicators. This integral focus is required to obtain performance information with the right qualities.

The options in the measurement system can be incompatible or complementary. In the former case, one option excludes another. In the latter case, the options can be cumulatively combined. This leads to a more extensive measurement system. A better integration of measurement processes may yield economies of scale. However, the incompatibilities in measurement design may be an obstacle. How can this problem be alleviated?

A first way is to integrate primarily measurement processes that serve the same functions. For instance, measurement systems for control and allocating resources (typically in the personnel department and financial department of the organization) may be integrated. A second option is to integrate primarily the step of the data collection since this usually the most costly phase in terms of money and administrative overhead. Thirdly, the problem can be overcome by investment. Organizations can fulfill the different functions of performance information by building an extensive measurement system with a good quality assurance system.

## **10.6. Conclusions from the fourth research chapter: What are the effects of performance measurement?**

Based on the literature, we listed seventeen effects of performance measurement. These effects are of two kinds. Seven effects concern measurement errors. The inputs, activities and outputs of organizations remain the same. Yet, the representation in numeric terms by means of measurement does not correspond with reality. Ten effects refer to change in the output of the organization. In this

case, inputs, activities and/or outputs are altered in order to comply with the indicators. We focused the remainder of the study on the latter effects of performance measurement.

Secondly, we looked for a common denominator for the effects. The concept of goal displacement is useful in this regard. Applied on the effects of performance measurement, this means that a good score on the indicators becomes the ultimate value. The fact that goal displacement may explain both intended and unintended effects is an important characteristic. It may bridge the gap between research findings of 'believers and disbelievers' in performance information.

Thirdly, we gave some empirical insight for the relative importance of the effects. The most likely effects have to do with increasing output of the measured activities (hypertrophy), often at the expense of attention for qualitative dimension (atrophy) or services that are not measured (tunnel vision). The effects that affect the production process (ossification), the intake (cherry picking) and the output levels (gaming) are moderately likely. The effects that result from poor target setting are the less probable ones. Myopia refers to short-term targets without a long-term view. In case of sub-optimization, targets are not societal optimal. Complacency results from targets that are not ambitious enough. Finally, giving difficult cases less priority is the least expected effect.

Fourthly, we linked the effect-profile of the twelve sections with the profiles of use. In general, more use leads to more effects. Accountability use does provoke the same level of effects as the use for management purposes. We found that the increase in effects of performance information could not be attributed to the use for research and learning. The latter use is distributed over the range of effects profiles.

One conclusion might be that use of performance information for research and learning is the most preferable for it has the least effects. Yet, it is the least developed in the sections of Ministry of the Flemish Community. The results of these uses are the least tangible. As a result, research and learning may also be the preferential use for symbolic reasons. There may be a thin line between use for real research and learning and the non-use of the performance information. Since accountability and management do not seem to be incompatible with use for research and learning, a multi track development is feasible.



## 11. Prospective directions

This study addressed several issues of performance measurement at a middle range level. We found amongst others that supply and demand need to fit, that measurability is an important factor in measurement development, that dynamics in policy sectors are substantially different, that measurement is contingent upon the societal environment, that performance measurement has effects of two kinds (altering output and manipulating measurement), and that they are related to the use of performance information. Middle Range Theories express the idea that social science – in our case Public Administration – should aim at explanations specifically tailored to a limited range of phenomena (Hedstrom and Swedberg 1998). Middle range theorizing does not in advance take on broad and abstract topics or try to establish universal laws. Yet, it has the ambition of integrating empirical findings and generalizations and to develop cumulative knowledge. Based on our experiences with middle range theories and empirical results in this study, we believe that the study of performance measurement in the public sector can be advanced in three ways.

1. by combining theories of organization with theories on information
  - a. extension of the view of organization
  - b. extension of the view of information
2. by recognizing more explicitly differences between policy sectors, and in particular
  - a. the role of information in policy processes
  - b. the tension between functional policies and administrative policies
3. by looking at reform theories to identify the main variables that explain dynamics, and identify the social mechanisms between macro trends and micro decisions.

### 11.1. A combination of theories of information and theories on organization

Theorizing about performance measurement in the public sector organizations is mainly inspired by institutional theory (e.g. Berry, Brower and Flowers 2000 Lawton, McKeivitt, and Millar 2000). Typical generalizations refer to the incentive structure underlying performance management (e.g. Dawson and Street 2000; Greener 2001), institutional barriers (e.g. Ammons 1985), symbolism (e.g. Modell 2004), organizational learning (e.g. Van Thiel and Leeuw 2002) and isomorphism (Roy and Seguin 2000). A step forward in the theoretical and conceptual development of performance measurement in public sector is an integration with theories of information utilization (e.g. DeLancer and Holzer 2001). Performance measurement in public sector organizations refers to two phenomena: information and organization. The study of performance measurement information therefore may advance by combining a theory of organization (usually of an institutional angle) with a theory of information.

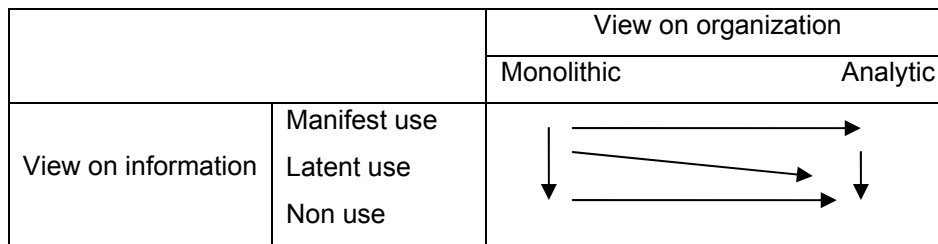


Figure 42: Prospective directions for combining theories on organization and information

Figure 42 represents some potential extensions of the research agenda. First, organizations can be viewed as monolithic units of analysis. Alternatively, an analytic view dissects the organization in different components. The components may be structural elements such as organizational subdivisions and organizational functions or cultural elements such as attitudes of employees. Secondly, the perspective on information may be limited to the manifest use of information. In addition, the latent uses may be taken into consideration. It is also conceivable to study the non use of information. We briefly discuss the two extensions<sup>93</sup>.

#### 11.1.1. Extension of the view of organizations.

In our study, organizations –sections of the Ministry of the Flemish Community- were the unit of analysis. In this sense, they were assumed to be monolithic decision making units. Characteristics were attributed to the organization as a whole. A useful extension would be to get inside the organization. Some organizational theories are better at explaining intra-organizational dynamics than others. Population-ecology for instance looks at organizations at the level of a population, and explains life and death of organizations mainly by population density (Hannan and Freeman 1977; Donaldson 1995). Clearly, such a perspective does not yield insight into organizational dynamics. We do not intend to review organizational theory here. We limit ourselves to presenting a theory for studying organizational dynamics. Brunsson's theory of talk, decision and action in organizations may be such an analytic theory of organization (2002).

The basic assumption of Brunsson's theory is that faced with multiple and conflicting societal demands, organizations will talk in one way, decide in another and act in a third (xiii)<sup>94</sup>. Organizations will meet

<sup>93</sup> These two extensions should be viewed as potential new avenues for research. They are seen as complementary perspectives to the perspective of this study. We do not consider one perspective to be superior.

<sup>94</sup> At first glance, the title of the book 'organization of hypocrisy' might suggest a skeptical view on organization. Yet, in the preface to the second edition, Brunsson makes clear that 'hypocrisy is seen as a solution rather than a problem, it possesses some moral advantages, and is often impossible to avoid (xi)'

some demands by way of talk, others by decisions and yet others by action. Some organizations predominantly produce talk (i.e. ideas). They are political and typically institutionalize conflict. Other organizations mainly produce action (i.e. products and services). They typically reduce conflict, and the need for decision-making through strong ideologies.

Ideas (talk) and actions are related to each other in four ways. First, a link can be absent. The ideologies of the ideologists or the plans of the planners may have little impact on the actions of the actors and vice versa. The ratings of US policy programs with the Program Assessment and Rating Tool for instance have only a modest impact on the budget requests (Norcross, 2005). The second configuration is the notion of ideas as control. This is the traditional normative model that assumes that ideas dictate actions. Ideas precede action. Thirdly, ideas may be explanations. This is the legitimating model that sees ideas as a post hoc explanation of instances that already occurred. Fourthly, ideas may be compensation. This is the compensating model. Organizational talk is adapted to some external demands or norms, and action to others. This is for instance the case when management talk protects the actions by satisfying the demands the action does not meet. Thus, ideas and actions are more or less considered to be coupled. Decisions are a means of coupling ideas and actions. In contrast to the rational model, talk and decision have a value in their own right and therefore are a valid empirical focus. For studying the relation 'of' or 'between'? performance information and organizational dynamics, this theory is valuable for at least two reasons.

First, talk, decisions and action have a value in their own right. Talk, decision and action are three types of output of organizations. They are not seen as necessarily subsequent steps in a process, where talk leads to decisions and decisions lead to action. Organizations can produce divergence between talk decisions and actions because they are confronted with inconsistent demands from the environment. They not only have to provide good products and services, they also have to comply with institutional structures, processes and ideologies. A school for instance is only to a limited extent assessed based on its product - teaching new knowledge. Parenthetically, it is difficult to judge whether pupils acquire new knowledge. In addition, it has to comply with curricula, pedagogical concepts, hire trained staff, etc. In some cases, the institutional rules are not the most efficient ones for delivering products. In order to meet these inconsistent norms, organizations may talk in one way, but talk in another way. Because of the nature of public service delivery, the institutional norms are more pressing in the public sector than in the private sector. The extent to which a public organization is satisfying the external demand by production is less observable compared to private organizations that work in markets.

Secondly, the framework can be applied at different levels of organization, from micro to macro. The same mechanisms may play in subdivisions in relation to the demands of the top of the organization. Subdivisions may talk in one way and act in another to satisfy demands of the top management. A distinction between talk decisions and action may also be useful at the macro level. The role of a

politician is mainly to produce talk and decisions. Parliaments for instance are deliberately structured for this purpose. Agencies usually have a to act and will thus need different kinds of information.

### 11.1.2. Extensions in the view of information: Knowledge Utilization.

One way to study the use of information is to focus on the manifest uses. That is what we did in our study. We looked at use for accountability, internal management and research and learning. Yet, once the information is provided, there are other possibilities. Information may be used in a latent way or may even not be used. The effects of not using information or using it in a latent way are a complementary field of study to the study of the effects of the manifest use of performance information. Knowledge Utilization theory has a longer tradition in studying different instances of use.

The professional field of knowledge utilization explores the strategies to put knowledge to use by both individuals and organizations (Backer 1991)<sup>95</sup>. Four theses serve as operating assumptions for the knowledge utilization field (Dunn and Holzner 1988). First, the *subjectivity thesis* states that knowledge is subjectively consumed by both organizations and individuals. The objective -read empirical- quality of the knowledge does not rule out this subjective dimension. Secondly, the *corrigibility thesis* implies that knowledge can be refined and improved. This is however a complicated undertaking. In fact, the problems and possibilities for improving knowledge are the core of a lot of knowledge utilization research. Thirdly, the *sociality thesis* involves that production, transfer and utilization of knowledge are inherently social processes, making social sciences relevant for the knowledge utilization field. Fourthly, the *complexity thesis* means that processes of knowledge creation, diffusion and utilization are interdependent in their causes and effects and thus complicated.

---

<sup>95</sup> The main journal in the field is titled *Knowledge: Creation Diffusion and Utilization (1980)*. Yet, there has traditionally been more attention for diffusion and utilization than for creation. Three waves of knowledge research may be identified (Backer 1991). The first wave extended from 1920 to 1960. The main issues were the diffusion of agricultural and educational innovations (Rogers 1983). Between 1960 and 1980, a second wave of interest and activity added two topics to the knowledge utilization arena. First, the focus on individual adoption of innovation was extended to organizational adoption. Secondly, the dissemination and innovations emerging from research and demonstration activities were studied. In the 1980s under the Reagan administration, the number of programs was sharply reduced. The third wave started in the early nineties. Again, two changes occurred (NCDDR 1996). First, the communications infrastructure of knowledge utilization has changed significantly (Paisly 1993: p. 222). Secondly, understanding of the process of knowledge utilization has shifted. The complexities and the dynamic, transactional aspects of knowledge utilization became subject of study (NCDDR 1996). Knowledge is not a static object to be sent and received, but a fluid set of understandings shaped both by those who originate it and by those who use it.



Although the link between knowledge utilization theory and performance measurement research is obvious, integrative efforts are rare. Innes (1990) pointed in the preface to a second edition of her 1975 book to the parallels between her work on institutionalizing of policy indicators and the work of Weiss on the uses of social research for policy-making. DeLancer and Holzer (2001) applied the research utilization model of Beyer and Trice (1982) - in particular, the distinction between adoption and implementation - to performance measurement. In our study, we made an analogous distinction between adoption and implementation.

Performance information is one kind of knowledge, which needs to be created, disseminated and used. The supply/production and demand/use analogy of performance information is fundamental to knowledge utilization theory too. The main addition of knowledge utilization theory to performance measurement research thus may be a refinement of the categories of use and users. We briefly describe some influential classifications, which may be relevant for performance information use too and which may give in particular some insight into in the latent 'grey' use of performance information. We briefly look at the work of Carole Weiss, Beyer and Trice and Innes.

At the end of the 1970s, Weiss published some influential articles about research utilization (Weiss 1977; 1979). It was a time of controversy about the impact of research on decision-making. After a period of expansion of research programs, many were disappointed by the incapability of research findings to impact and guide policy decisions (Williams 1998). Weiss showed that expecting direct impact may be unrealistic. Yet, there are several other ways in which research may have impact. One of her best known concepts is knowledge creep<sup>96</sup>. Research gradually - or in some cases quickly - spreads, enters into use, and sometimes becomes the conceptual framework of entire policy debates. Although the impact of research results cannot directly be pinpointed to a specific decision, the parameters of the discussion may be set by research<sup>97</sup>.

Weiss thus found that research has a broader array of functions than generally assumed. The traditional rational model of research that guides decisions and solves problems had to be expanded

---

<sup>96</sup> An example of knowledge creep is the discovery of the Matthew effect which essentially states that those with the best chances (in terms of socio-economic background) in society often benefit the most from the provision of public goods such as education and social security. Although the observation of this phenomenon probably did not influence decisions directly, it had an important impact in framing the policy debate about the welfare state in Belgium (Deleeck et al. 1983). Indicators may also have this potential.

<sup>97</sup> Weiss' theory is a positive one. The normative stand in her work is the thesis that more use of research knowledge in public policy evaluation will increase the quality of policies. She states on evaluation that "even if we realize that evaluation is not the star in the policy drama, we have a

in order to account for other roles. Weiss distinguished between six models of research utilization (Table 45). These six meanings of research utilization relate to different functions of research knowledge Weiss (1979). The first two models, the knowledge driven model and the problem solving model are the more generally accepted uses. They represent a stimulus-response type of process where research directly stimulates innovation (for the knowledge driven model) and problem solving (for the problem solving model). The four other models point to more latent functions. The interactive model adds that the process between research and decision is not linear. Researchers are also involved in making sense of a problem. The political model points to research which function is support rather than illumination. Predetermined stands are confirmed. The tactical model is about research that plays a role in political strategy, for instance to delay a decision without giving the impression of immobility. Finally, the enlightenment model refers to research that sets the stage of policy debates. The vocabulary, the solutions and problems of talking about a policy issue are influenced by research results. Knowledge creep is an expression of this category.

<i>model</i>	<i>Definition</i>
knowledge driven	basic research -> applied research -> development -> application
problem solving	problem -> existing or new research -> interpretation -> policy choice
interactive	issue areas that pool talents, beliefs and understandings
political	stand from interest, ideology or intellect -> research support
tactical	research as a maneuver, substance of research is less relevant
enlightenment	research diffuses in society -> defines problem definition and solution

Table 45: Weiss' (1979) models of research utilization

Like Weiss, Beyer and Trice asked the question why research is not used as extensively and as often hoped for. They also stress that the prevalent foci of research - the attitudes of potential users toward research and whether or not research is used in making decisions - need to be extended to other utilizing actions. However, unlike previous authors, they approach utilization as a process, rather than a condition. The process consists of two main phases; the adoption phase and the implementation phase. Each phase consists of different specific behaviors which can be organized in both rational and less rational sequences. In order to study the utilization process, they draw heavily on organization theory. The process approach to knowledge utilization is the main attribute of Beyer and Trice's work.

Three different uses are defined (Beyer and Trice 1982: p598). First, instrumental use involves acting on research results in specific direct ways. Secondly, conceptual use involves using research results for general enlightenment. Results influence actions, but in less specific, more indirect ways. Thirdly, symbolic use involves using results to legitimate and sustain predetermined positions. These uses are reflected in the processes. Different uses will lead to different processes. Conceptual use for instance

---

*responsibility to communicate the best information and analysis available to the principal players."* (Weiss, 1999, p. 483).

gives great latitude to users in selecting, redefining, altering, combining and generally reinterpreting research results to fit a wide variety of circumstances (p.600).

Innes's book 'knowledge and public policy' applies knowledge utilization on indicators. The knowledge utilization was the implicit perspective in the 1975 edition. The second expanded edition made the link more explicit. Concepts were reframed in order to fit within the knowledge utilization school. The study is based on an historical account of three indicators. The unemployment rate is a successful indicator because it is widely used and highly institutionalized. The standard budget is another indicator that is widely used. However, it is and indicator without a theory. It attempts to measure a poorly specified concept - the societal norm for the level at which people can or should live. The crime rate is seen as an inadequate indicator mainly because of unsatisfactory validity or reliability.

Several functions of knowledge are found throughout the text. First, indicators have a role in the identification of problems. Indicators such as the unemployment or the inflation rate enable the definition of a problem. Unemployment figures higher than around or higher than 10% are problematic. An inflation rate of 2% is considered normal, more is problematic. The problem-defining role of indicators is also described as a pre-policy use. Secondly, indicators play a role in policy itself. The unemployment figures are an example of an indicator for policy - for norm setting and for causal analyses. Thirdly, indicators play a role in administration. The standard budget for instance was for a long time only used by the administration to calculate wages and benefits. Fourthly, indicators are used for program design. They are integrated in policy instruments such as subsidies and contracts.

A further refinement would be to make a distinction between three aspects of organizational life: talk decisions and actions. The cross tabulation of use and talk decision and action would give a detailed picture of the use of information in the organization. The use (and non-use) of information will feed back to the supply and demand function. Non-use for instance may be countered by more or less demand for measurement. More, because measurement is seen as insufficient. Less, if measurement is seen as redundant. Yet, as we discussed in chapter 9, measurement has effects on the organization. These effects may be of three kinds: there may be manipulation of measurement, altering the output of the organization, or both. Manipulation of measurement is unintended. A change in the output may be intended or not intended. Again, experiences with measurement feed back into the supply and demand for information. For example, as a response to manipulation of information, organizations may attempt to build more robust measurement systems with heavier controls. Another option would be to reformulate demand towards less coercive uses of information<sup>98</sup>.

---

<sup>98</sup> Goddard Mannion and Smith (1999) for instance discuss the relation between soft (qualitative assessments) and hard data (quantitative measurement). They may be complements where qualitative information is used

## 11.2. The policy sector as an intermediate level of analysis

The study of performance measurement in public sector may benefit from a cross-sectoral angle. Studies usually focus on or the dynamics within a specific policy sector, or on administrative policies. A comparison of dynamics in policy sectors is more uncommon. In particular, the tension between administrative policies and dynamics in policy sectors are rare. Two useful models for mapping dynamics in policy sectors are the advocacy coalition framework (Sabatier 1988) and Benson's framework for policy analysis. Brunssons's theory is complementary to these approaches.

The principal ambition of Sabatier was to explain policy change. Therefore he formulated the Advocacy Coalition Framework (ACF). The main unit of analysis is the policy subsystem. Policy change occurs when external parameters alter. Some external parameters are more stable than others. Change in the external parameters is reflected in the constraints and resources of the policy subsystem. A policy subsystem consists of a multitude of actors that are involved in a policy issue. However, within a subsystem usually only two to four belief systems exist. These belief systems combined with a set of resources are fundamental for advocacy coalitions. The latter is a coalition of actors that develops a strategy to influence (a) decisions by sovereigns, (b) agency resources and general policy orientations, (c) policy outputs and ultimately and (d) policy outcomes. Policy brokers are go-betweens between the coalitions. The struggle between advocacy coalitions is ended when a sovereign takes a decision. Next, policies are implemented by agencies.

Sabatier's framework is consistent with Benson's view on policy sector as a level of analysis. Benson (1982) identifies two levels of structure in policy sectors. The first level includes the administrative structure, the policy paradigms and the resource dependencies. The administrative structure includes issues of differentiation and control such as the division of planning and implementation, expert or legal control and hierarchy or market control. The policy paradigm refers to the set of potential policy choices within a sector. Resource dependency is about the financial and political support for the organizations in the network. The second level according to Benson is the interest/power structure and the rules of structure formation. The interest/power structure are demand groups, support groups, administrative groups, provider groups, and coordinating groups. These are the actors in the advocacy coalitions according to Sabatier. The rules of structure formation should be seen as the boundaries to the actions of the actors in networks. There are two types of rules of structure formation, negative and

---

for digging into issues signaled by measurement. They may be substitutes where qualitative information is pushed away by quantitative information or vice versa. Finally, hard information may be a safety net when qualitative assessments fail. This will probably be the case for policy contracts. In times of conflict, the quantitative targets will come to the fore. When there is mutual understanding between principal and agent, they will mainly be symbolic.

positive ones. Positive selection rules require a sector to contribute to the reproduction of the social formation<sup>99</sup>. Negative selection rules define activities that are not permitted in organizations in the state because 'they violate its essential character as a capitalist state (p.162)'. The same boundaries of viable solutions are found in policy sectors.

Sabatiers advocacy coalitions framework has a link with the knowledge utilization school (see for instance Lindquist 1990). At the core of the advocacy coalition theory is the concept of a belief system. A belief system involves value priorities, perceptions of important causal relations, perceptions of world states (including the magnitude of a problem), perceptions of the efficacy of policy instruments, and so on. The structure of a believe system consists of three layers: a deep core of normative and ontological axioms, a near core of fundamental policy positions, and the secondary aspects such as instrumental decisions and information searches. Actors who share a belief system will more easily exchange and accept knowledge. Performance indicators play a role in belief system. Measurement results may challenge or confirm belief systems. On another level, some belief systems may be more tolerant towards measurement than others.

An important characteristic of a policy subsystem may be the degree of conflict or convergence between actors, both at the level of the belief system as the strategic and tactical positions actors take. Several potential conflicts can be identified. First, there may be a conflict between advocacy coalitions. Proponents of one coalition may be in conflict with another. These conflicts will have different intensity. Secondly, there may be conflict between the advocacy coalition and the delivery coalition. We use the term 'delivery coalition' for the network of public and private actors that implement policies. In essence, this is a conflict between the talk in a policy sector and the most important actions. These conflicting positions are relevant for performance measurement because they will impose different demands on public sector organizations that have to do the measuring. This may result in conflicting demands for information as well as conflicting definitions of performance.

An important issue for public sector organizations are conflicting demands from different policy subsystems. Public sector organizations are usually confronted with at least two policy subsystems. First there is the functional policy subsystem: education for the department of education, employment for a job placement agency. Secondly, there is the administrative policy subsystem. Administrative policy is about the organization of government. It mainly deals with horizontal issues such as personnel policy, the statute of the civil service, the budgeting procedures, format and functions, the accounting standards, the control pyramid including internal control, internal and external audit, the

---

<sup>99</sup> Benson cites Bourdieu's work (Bourdieu and Passeron 1977) to explain the positive selection rules. Bourdieu's studies social reproduction: how does one generation of a group makes sure that it reproduces itself and passes its qualities to the next.

statistical system, the organizational chart of the ministries, the rules for contracting, etcetera. Demands for information may conflict between these subsystems.

A particular approach to studying performance information in policy subsystems might be to look at policy instrumentation - the design of policy instruments. For instance, in order to distribute subsidies and funds, governments need criteria. These criteria are almost always quantitative. In the case of repressive policy instruments, governments often need to define a threshold that triggers sanctioning action. Examples are emission standards for companies or minimal test scores for schools. Although these are not typical measurement initiatives according to most Public Administration scholars, these are significant uses of performance information – perhaps the most significant ones.

### **11.3. Accounting for macro level factors.**

A focus on policy sectors should not obfuscate the impact of macro level factors - government wide and society wide. Recent literature on public sector reform studied the global reform agenda that was embodied in the New Public Management. The central question was to what extent NPM took hold. Christensen and Laegreid (2002) distinguish three schools of thought in NPM implementation studies. The first school regards the implementation of NPM as a response to external pressure. The environmental determinism can be of two kinds. Symbolists would state that pressure arises from reform myths that exert isomorphic pressures – the pressure to align success stories as a way of dealing with uncertainty and as a substitute for success. Rationalists would attribute external pressure to the superior technical efficiency of the NPM approaches. Obviously, a symbolist would call this technical efficiency a myth. The second school of thought views NPM as a product of the path-dependent national historical-institutional context. National reforms have unique features. A successful implementation depends on the consistency of the underlying values of the reform with the existing values in the administrative system. The third school explains reform by the political-administrative structure. The main features of the polity, the form of government, and the formal structure of decision-making affect the country's capacity to realize administrative reforms. Christensen and Laegreid (2002) propose a transformative approach (2002). Institutional dynamics have to be understood as a mix of the factors outlined above: environmental factors (economy and ideology), the historical-institutional context and style of governance, and polity features. The transformation process will alter reform ideas, solutions and content, and implementation. As a result, countries develop local hybrid variants on the global NPM story.

An important issue thus is how macro factors filter through to the micro level and back (Hedstrom and Swedberg 1998). Pollitt and Bouckaert (2004) point to the importance of elites in the transformation of global trends to local versions. Elite-decision making (mainly by executive opticians and senior civil servants) is pivotal in public sector reform (Pollitt and Bouckaert 2004: p 26). Elites decide upon both desirability and feasibility of reform. Elites are the main transformers of reform ideas by confronting

them with macro factors such as the socio-economic situation. They will also consider what is feasible within the polity system, and if necessary and possible try to change the polity system. The mechanisms that explain the organizational predispositions towards performance measurement (macro→micro) and the mechanisms through which performance measurement at organizational level influences macro factors may be subject of more extensive research.

This study attempted to research empirically some aspects of performance measurement in public sector organizations. We hope to have demonstrated that a better match between supply and demand, combined with a realistic assessment of causes and effects is of value for this sustainable development of measurement. Additionally, we attempted to advance the theory formation on a middle range level. Some propositions were tested empirically. In the conclusion, we suggested some new courses of development. Future directions may be a better combination of theories of organization with theories of information, a better theorizing of the policy context of public sector organizations with special attention for the role of information in policy processes and the tension between functional policies and administrative policies, and a better insight of the mutual influence of the macro polity and societal context and the micro organizational context. We explicitly took a Public Administration perspective to study the functioning of measurement in public administrations. The research had as an explicit aim to further sustainable development of performance measurement in public administrations. Herewith, we also attempted to meet the practical challenge and relevance of the study of the subject, which is the development of sustainable performance measurement systems - to find what public managers and politicians can *reasonably expect from it*.





## 12. References

- Abravanel (2004) *Surveying clients about outcomes*. Washington DC: The Urban Institute.
- Administratie Planning en Statistiek (2003) *Kwaliteitszorg Statistisch Productieproce: Aanbevelingen*. Brussel APS
- Adnum D. (1993) Using performance indicators for effective public sector management. *Management Accounting*. 71:1. pp.48-50.
- Alchian, A., and Demsetz, H. (1972). Production, Information Costs and Economic Organization. *American Economic Review*. 62:5. pp.777-795.
- Alcock, P. (2004) Targets, indicators and milestones: what is driving area based policy in England? *Public Management Review*. 6:2. pp.211-227.
- Allen, W.H. (1906). Hospital Efficiency. *The American Journal of Sociology*. 12:3. pp.298-318.
- Ammons, D. N. and Rodriguez, A. (1986). Performance appraisal for upper management in city governments. *Public Administration Review*. 46:5. pp. 460-467.
- Ammons, D. N., Coe, C. C. and Lombardo, M. (2001). Performance comparison projects in local government: participants' perspectives. *Public Administration Review* , 61:1. pp.100-110.
- Ammons, D.N. (1985). Common barriers to productivity Improvement in local government. *Public Productivity Review*. 9. pp. 187-202.
- Ammons, D.N. (1995). *Accountability for performance: measurement and monitoring in local government*. Washington: ICMA:
- Ammons, D.N. (2002) *Tools for decision making: a practical guide for local government*. Washington: CQ press.
- Ammons, D.N. (2002). Performance and managerial thinking. *Public Performance and Management Review*. 25:4. pp.344-347.
- Amsden, R.T., Ferratt, T.W., and Amsden, D.M. (2001). TQM: Core Paradigm Changes. in: Stupak, R. J. and Leitner, P. M. *Handbook of public quality management*. New York: Dekker. pp. 129-741.
- Anderson, K. H. Burkhauser, R. V. and Raymond J. E. (1993): The Effect of Creaming on Placement Rates Under the Job Training Partnership Act. *Industrial and Labor Relations Review*, 46:4. pp.613-624.

- Anderson, M. J. and Fienberg, S. E. (2001). *Who counts? The politics of census-taking in contemporary America*. New York: Russell Sage foundation.
- Andreoni, J. (1990) Impure Altruism and Donation to Public Goods: A theory of warm-Glow Giving. *Economic Journal*, 100. pp.467-77.
- Andrews, F.M. (1989). The Evolution of a Movement. *Journal of Public Policy*. 9:4. pp.401-405.
- Andrews, R. (2004). Analysing Deprivation and Local Authority Performance: The Implications for CPA. *Public Money and Management*. 24:1. pp.19-26.
- Andrews, R., Boyne, G.A., Meier, K.J., O'Toole, Jr.L.J., and Walker, R.M. (2005) Representative Bureaucracy, Organizational Strategy, and Public Service Performance: An Empirical Analysis of English Local Government. *Journal of Public Administration Research and Theory* 15:4. pp.489-504.
- Anthony R.N. (1965) *Planning and control systems; a framework for analysis*. Boston: Graduate School of Business.
- Aristigueta, M.P. (2000). The Rise and Fall of the Florida Benchmarks. In: A. Neely, *Performance Measurement-Past, Present, and Future*. Cambridge: Fieldfare Publications Ltd. pp.16-23.
- Aristigueta, M.P. (2004). Indicators for Living conditions: Challenges and Opportunities for Accountability and Governance. in: Holzer, M. and Lee S-H.. *Public Productivity Handbook*. Second Edition, Revised and Expanded. New York (NY): Marcel Dekker. pp. 431-447.
- Aristigueta, M.P., Cooksy, L.J., and Nelson, C.W. (2001). The Role of Social Indicators in Developing a Managing for Results System. *Public Performance and Management Review*. 24:3. pp. 254-269.
- Arnstein, S.R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners*, 8:3. pp.217-224.
- Asian Development Bank (2004) *Governance: Progress and challenges in Mongolia*. Manila. ADB.
- ASPA. Center for accountability and performance. (2000). *Performance measurement: concepts and techniques*. Washington: American society for public administration.
- Aucoin, P. and Heintzman, R. (2000) The dialectics of accountability for performance in public management reform. *International Review of Administrative Sciences*. 66:1. pp.45-55.
- Babbie, E.R. (1975) *The Practice of Social Research*. Belmont: Wadsworth.
- Backer, T. (1991). Knowledge utilization: The third wave. *Knowledge*. 12:3. pp.225-240.
- Barrett, K. and Greene, R. (2005) Grading the States: a management report card. *Governing*. 18:5. pp. 24-34.

Barzelay, M. (1992) *Breaking Through Bureaucracy: a new vision for managing in government*. University of California Press: Berkeley.

Bauer, R.A. (ed.) (1966). *Social Indicators*. Cambridge: MIT Press.

Behn (2001). *Rethinking Democratic Accountability*. Washington, DC: Brookings Institution.

Behn, R. D. (2003) Why measure performance? Different purposes require different measures. *Public Administration Review*. 63:5. 586-606.

Behn, R. D. And Kant, P. A. (1999). Strategies for avoiding the pitfalls of performance contracting. *Public Productivity and Management Review*. 22:4. pp. 470-489.

Behn, R.D. (2002). The psychological barriers to performance measurement: Or why isn't everyone jumping on the performance management bandwagon? *Public Performance and Management Review*. 26:1. pp.5-25.

Bell, J. (1993). *Doing Your Research Project*. Buckingham: Open University Press

Ben-Chieh, L. and Livingston, R. C. (1975). *Quality of life indicators in U. S. metropolitan areas, 1970: a comprehensive assessment*. Washington: Environmental Research Center.

Benson, J.K. (1982). A Framework for Policy Analysis. In D.L. Rogers and Whetten D.A. (eds.). *Interorganizational Coordination: Theory, Research, and Implementation*. Ames: Iowa State University Press. Pp.137-176

Berelson, B. (1952) *Content Analysis in Communication Research*. New York: Hafner Publishing Company.

Berliner, J. S (1956) A Problem in Soviet Business Administration. *Administrative Science Quarterly*, 1:1. pp.86-102.

Berman, E. (2002). How Useful is performance management? *Public Performance and Management Review*. 25:4. pp.348-351.

Berman, E. and Wang, X. (2000). Performance measurement in U.S. Counties: Capacity for reform. *Public Administration Review*. 60:5. pp.409-20.

Berry, A.J., Broadbent, J. and Otley, D.T. (1995) *Management control: theories, issues, and practices*. Houndmills, Basingstoke, Hampshire: Macmillan

Berry, F. S., Brower, R. S. and Flowers, G. (2000). Implementing performance accountability in Florida: what changed, what mattered, what resulted? *Public Productivity and Management Review*. 23:3. pp.338-358.

- Best, J. (2004) *More Damned Lies and Statistics*. Berkely: University of California Press.
- Beyer, J.M. and Trice, H.M. (1982). The utilization process: A conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly*, 27 pp. 591-622.
- Black, N. (2001). Evidence Based Policy; Proceed with Care. *British Medical Journal*. 323. pp. 275-279.
- Blankart, C.B. (1987) *Limits to privatization*. *European Economic review*. 31. pp.346-351.
- Blau, P. M. (1955). *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies*. Chicago: University of Chicago Press
- Blau, P. M. and Scott, R. (1962). *Formal Organizations: A Comparative Approach*. San Francisco: Chandler.
- Boaz, A. and Nutley S.M. (2003) Evidence-based Policy and Practice. In; T Bovaird and E Loffler (eds) *Public Management and Governance*. London: Taylor and Francis
- Bobrow, D.B. and Dryzek, J.S. (1987). *Policy analysis by design*. Pittsburgh, Pa.: University of Pittsburgh Press
- Bockstaele, P., Cerulus, F. and Vanpaemel, G. (2004). *Ars Conjectandi: Over gokkers, geleerden en grote getallen (Ars Conjectandi: about gamblers; scientists and big numbers)*. Documentation of the Exhibition at the University of Leuven. 26th of May - 27th June 2004
- Bohte, J. and Meier, K.M. (2000) Goal displacement: assessing the motivation for organizational cheating. *Public Administration Review*. 60:2. pp.173-182
- Bolton, M. (2003) Public sector performance measurement: delivering greater accountability, *Work Study* 52.1: 20-24.
- Bouckaert, G. (1990). The history of the productivity Movement. *Public Productivity and Management Review*. pp. 14:1.
- Bouckaert, G. (1992), Productivity analysis in the Public Sector: the case of Fire service. *International Review of Administrative Sciences*. 58:2. pp175-200.
- Bouckaert, G. (1992). Public productivity in retrospective. In: Holzer, M. (Ed.). *public productivity handbook*, New York: Marcel Dekker, pp. 15-46.
- Bouckaert, G. (1993) Measurement and meaningful management. *Public Productivity and Management Review*. 17:1. pp.31-43.

- Bouckaert, G. (1993), Efficiency measurement from a management perspective: a case of the civil registry office in Flanders. *International Review of Administrative Sciences*. 59:1. Pp.11-27.
- Bouckaert, G. (1995). Improving Performance Measurement. In: Halachmi & Bouckaert (eds). *The Enduring Challenges of Public Administration*. San Francisco. Jossey Bass. pp. 379-412.
- Bouckaert, G. (2004). Institutionalising Monitoring and Measurement Systems in the Public Sector. In: Benz, A, Siedentopf, H. and Sommermann, K-P. *Institutionwandel in Regierung und Verwaltung: Festschrift für Klaus König zum 70. Geburtstag*. Berlin; Duncker & Humblot.
- Bouckaert, G. and Auwers T. (1999a) *De modernisering van de Vlaamse Overheid (The modernization of the Flemish government)*. Overheidsmanagement nr. 2. Brugge: die Keure.
- Bouckaert, G. and Auwers T. (1999b) *Prestaties meten in de overheid. (Performance measurement in Government)*. Overheidsmanagement nr. 5. Brugge: die Keure.
- Bouckaert, G. and Balk, W. (1991). Public productivity measurement: Diseases and cures. *Public Productivity & Management Review*, 15:2. pp.229-235.
- Bouckaert, G. and Halachmi, A. (1995) The range of performance indicators in the public sector: theory versus practice. In Halachmi, A. and Grant, D. *Reengineering and Performance Measurement in Criminal Justice Social Programmes*. Perth: IIAS. Pp. 91-106.
- Bouckaert, G. and Peters, B.G (2002). Performance Measurement and Management. The Achilles' Heel in administrative modernization. *Public Performance and Management Review*. 25:4. pp.359–362.
- Bouckaert, G. and Thijs, N. (2003). *Kwaliteit in de overheid (Quality in Government)* Leuven: Instituut voor de overheid.
- Bouckaert, G. and Van Dooren, W. (2000) *Subsidiëren van Nederlands-Vlaamse samenwerking: hindernissen en oplossingen*. Brussel: Commissie Cultureel Verdrag Vlaanderen-Nederland
- Bouckaert, G., Hoet D. and Ulens W. (2000) *Prestatiemeetsystemen in de overheid: Een internationale vergelijking. (Performance Measurement systems in Government: an international comparison)* Overheidsmanagement nr. 9. Brugge: die Keure.
- Bouckaert, G., Van Dooren, W. and Sterck, M. (2003). *Prestaties Meten in de Vlaamse Overheid: een verkennende studie (Performance measurement in Flemish government: an explanatory study)*. Leuven: SBOV.
- Bouckaert, G., Van Reeth, W., Auwers, T. and Verhoest K. (1998). *Handboek doelmatigheidsanalyse & prestaties begroten* (handbook efficiency analysis and performance budgeting). Brussel: Ministerie van de Vlaamse Gemeenschap

- Bourdieu, P. and Passeron, J-C (1977) *Reproduction in education, society and culture*. SAGE studies in social and educational change. Beverly Hills: Sage.
- Bovaird, T. and Loffler, E. (eds) (2003). *Public Management and Governance*, London: Taylor and Francis
- Bowerman, M. (1995) Auditing performance indicators: The role of the Audit Commission in the Citizen's Charter Initiative. *Financial Accountability and Management*. 11:2. pp.171-183.
- Bowles, J. and Hammond, J. (2001). TQM apostles. in: Stupak, R. J. and Leitner, P. M. *Handbook of public quality management*. New York: Dekker. pp. 40-44
- Box, R. C., Marshall, G. S., Reed, B.J. and Reed, C. M. (2001). New Public Management and Substantive Democracy. *Public Administration Review*, 61:5, pp608-620.
- Boyne G. and Walker, R.M. (2005) Introducing the "Determinants of Performance in Public Organizations" Symposium. *Journal of Public Administration Research and Theory* 15:4. pp.483-488
- Boyne, G. (2004) Explaining Public Service Performance. Does Management Matter? *Public Policy and Administration*. 19:4. pp.110-117.
- Boyne, G. and Law, J. (1991). Accountability and local authority annual reports the case of the Welsh District councils. *Financial Accountability and Management*. 7:3. 179-94.
- Boyne, G.A. and Gould-Williams, J. (2003) Strategic Planning and Organizational Performance in Public Organizations *Public Management Review* 5:1.
- Bradley, S., Crouchley, R, Millington, J. and Taylor, J. (2000) Testing for quasi market forces in secondary education. *Oxford Bulletin of Economics & Statistics*. 62:3 pp.357-391.
- Brans, M. (2003) Comparative Public Administration: From General Theory to General Frameworks. in Peters, B.G. and Pierre, J. (ed.) *Handbook of Public Administration*. New York: Sage.
- Brans, M. and Rossbach S. (1997) The Autopoiesis of Administrative Systems: Niklas Luhmann on Public Administration and Public Policy. *Public administration* 75:3. pp.417-440.
- Brewer G.A. and Selden S.C. (2000). Why Elephants Gallop: Assessing and Predicting Organizational Performance in Federal Agencies. *Journal of Public Administration Research and Theory* 10:4. pp.685-711.
- Brewer, G.A. (2005) In the Eye of the Storm: Frontline Supervisors and Federal Agency Performance. *Journal of Public Administration Research and Theory* 15:4. pp.505-527.
- Broom, C.A. (1995). Performance-based government models: building a track-record. *Public Budgeting and Finance*. 15:4. pp.3-17.

- Brown, R.E. and Pyers, J.B. (1988). Putting teeth in efficiency and effectiveness of public services. *Public Administration Review*. 48:3. pp.735-742
- Brunsson, N. (2002). *The Organization of Hypocrisy: Talk, Decisions and Actions in Organizations*. Wiley: Chichester.
- Buchanan, J.M., and Tullock G. (1962) *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.
- Budge, I., Robertson, D. and Hearl D. (eds.) (1987). *Ideology, Strategy and Party Change: Spatial Analysis of Post-War Election Programs in 19 Democracies*, Cambridge: Cambridge University Press.
- Bulmer, M. (2001). Social Measurement: what stands in its way? *Social Research*. 68:2. pp. 455-480
- Bulmer, M., Bales K., and Sklar K.K (eds.) (1991). *The Social survey in historical perspective, 1880-1940*, Cambridge: Cambridge University Press.
- Caiden, G. E. (1991). What Really is Public Maladministration? *Public Administration Review* 51:6. pp.486-493.
- Caiden, N. (1981) Public Budgeting amidst Uncertainty and Instability. *Public Budgeting and Finance*, 1:1. pp.6-19.
- Carley, M. (1981). *Social Measurement and Social Indicators: Issues of Policy and Theory*. London: George Allan and Unwin.
- Carter, N. (1991). Learning to measure performance: the use of indicators in organizations. *Public Administration*. 69:1. pp.85-101.
- Cavalluzzo, K. and Ittner, C. (2004) Implementing Performance Measurement Innovations: Evidence from Government. *Accounting Organizations and Society*. 29:314. pp.243-267.
- Cazes, B. (1974). The Development of Social Indicators: a Survey. in Schonfield, A. (ed.) *Social indicators and social policy*. London: Heinemann.
- Chalos, P. and Cherian, J. (1995). An application of Data Envelopment Analysis to public sector performance measurement and accountability. *Journal of Accounting and Public Policy*. 14:2. pp.143-160
- Cherchye L, Moesen W, Van Puyenbroeck T, (2004) Legitimately diverse, yet comparable: Synthesising social inclusion performance in the EU *Journal of Common Market Studies*, 42:5. pp.919-955.
- Christensen, T. and Laegreid P. (eds.) (2002) *New public management: the transformation of ideas and practice*. Aldershot: Hampshire.

Christensen, T. and Laegreid, P. (2001). New Public Management: the Effects of Contractualism and Devolution on Political Control. *Public Management Review*, 3:1. pp.73-94.

Christensen, T., Laegreid, P. and Stigen, M. (2004) *Performance Management and Public Sector Reform: The Norwegian Hospital Reform*. Paper presented at the EGPA conference "Four Months After: Administering the New Europe" - Study Group on Productivity and Quality in the Public Sector, "Performance Measurement and Management in the Public Sector", Ljubljana, September 1-4, 2004.

Christensen, T., Laegreid, P. and Wise, L.R. (2002). Transforming Administrative Policy. *Public Administration*. 80:1. pp.153-178.

Chun, Y.H. and Rainey H. G. (2005) Goal Ambiguity and Organizational Performance in U.S. Federal Agencies. *Journal of Public Administration Research and Theory*. 15:4. pp.529-557.

Churchman, C. W. and Schainblatt, A. H. (1969). PPB: how can it be implemented? *Public Administration Review*, 29:2, pp.178-190.

Cobb, C.W. and Rixfors, C. (1998). *Lessons Learned from the History of Social Indicators*. San Francisco: Redefining Progress.

Coe, C. K. (2004). A report card on report cards. *Public Performance and Management Review*. 27:2. pp.53-76.

Conings, V., Sterck, M, Van Dooren, W., and Bouckaert, G. (2005) *Beleidsnota's versus BBB: Een toetsing van de beleidsnota's aan het Comptabiliteitsdecreet en het Kaderdecreet Bestuurlijk Beleid*. Leuven: SBOV.

Council of the European Union (2005). *European Council Brussels 22 and 23 March 2005: Presidency Conclusions*. Brussels: Council of the European Union.

Courty, P. Marschke, G. (2003) Performance Funding in Federal Agencies: A Case Study of a Federal Job Training Program. *Public Budgeting & Finance*. 23:3. pp.22-49.

Crosby, P. (1979). *Quality is Free - The Art of Making Quality Certain*. New York: McGraw Hill.

Davidson, R. (1991). The Social Survey in historical perspective: a governmental perspective. in: Bulmer, M., Bales K., and Sklar K.K (eds.). *The Social survey in historical perspective, 1880-1940*, Cambridge University Press. pp.359-368.

Davies, H. T. O. & Nutley, S. M. (2000) Developing learning organisations in the new NHS, *British Medical Journal*, 320, pp. 998-1001.

Davies, H. T. O., Nutley, S. M. & Smith, P. C. (eds) (2000) *What Works? Evidence-Based Policy and Practice in Public Services*. Bristol: The Policy Press.



- Davies, I.C. (1999). Evaluation and Performance Management in Government. *Evaluation*. 5:2. pp.150-159.
- Davies, P. (2000). What is Evidence Based Education? *British Journal of Educational Studies*. 47:2. pp. 108-121
- Dawson, D. and Street, A. (2000). Comparing NHS Hospital Unit Costs. *Public Money and Management* 20:4 pp.58-62.
- De Borger B., Kestens K., Moesen W. and J. Vanneste (1994) Explaining differences in productive efficiency: an application to Belgian municipalities, *Public Choice*, 80:3/4 pp.339-358.
- De Peuter, B., Bouckaert, G. and Van Dooren, W. (2003). *Metten en vergelijken van lokale bestuurlijke ontwikkeling: een monitoringsysteem voor het lokale bestuur in Vlaanderen. (Measuring and Comparing local administrative capacity: a monitor for local government in Flanders.)* Brugge: Die Keure.
- Deci, E. L. (1975). *Intrinsic Motivation*. New York: Plenum.
- DeLancer Julnes, P. (2001). Does participation increase perceptions of usefulness?: An evaluation of a participatory approach to the development of performance measures. *Public Performance and Management Review*. 24:4. pp.403-418.
- DeLancer Julnes, P. and Holzer, M. (2001). Promoting utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review*. 61:6. pp.650-665.
- Deleeck, Cantillon, and Huybrechts (1983), *Het Matteuseffect: De ongelijke verdeling van de sociale overheidsuitgaven in België (The Matthew effect: the uneven distribution of social expenditures in Belgium)*. Antwerpen: Kluwer.
- Deming, W. (1986). *Out of the Crisis: quality, productivity and competitive position*. Cambridge: Cambridge University Press.
- Denhardt, R.B. (2000) *Theories of Public Organization (third edition)*. Fort Worth; Harcourt Brace College Publishers.
- Denhardt, R.B. and Denhardt, J.V. (2000). The New Public Service: Serving Rather than Steering. *Public Administration Review*, 60:6. pp.549-560.
- Denzin, N.K. (1970) *The research act: a theoretical introduction to sociological methods*. Chicago: Aldine.

- Denzin, N.K., (1978). Strategies of Multiple Triangulation. in Denzin N.K. (ed.) *The Research Act: a theoretical introduction to sociological methods*. Chicago: Aldine.
- Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring Labor Inefficiency in Post Offices, in Marchand, M., Pestieau, P. and Tulkens, H. (eds.) *The Performance of Public Enterprises: Concepts and Measurements*, pp.243-267.
- Desrosières, A. (1998). *The politics of large numbers: a history of statistical reasoning*. Cambridge: Harvard university press.
- DiMaggio, P. and Powell, W.W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48:2. pp.147-160.
- Donaldson, L. (1995). *American Anti-Management Theories of Organization*. Cambridge: Cambridge University Press.
- Downs, A. (1967) *Inside bureaucracy* Glenview, ill.: Scott, Foresman and Co.
- Dowrick and Quiggin (1998). Measures of Economic activity and Welfare: the use and abuses of GDP. in: Eckersley (ed.). *Measuring Progress*. Collingwood: CSIRO.
- Dubnick, M.J. (1998). Clarifying Accountability: an Ethical Theory Framework. In: Sampford, C. and Preston, C. (eds.) *Public Sector Ethics: Finding and Implementing Values*. London: Routledge.
- Dubnick, M.J. (1999). *Demons, Spirits, and Elephants: Reflections on the Failure of Public Administration Theory*. Prepared for delivery at the 1999 Annual Meeting of the American Political Science Association, Atlanta Hilton and Towers and Atlanta Marriott Marquis, September 2-5, 1999.
- Dubnick, M.J. (2005) Accountability and the Promise of Performance: In Search of the Mechanisms. *Public Performance and Management Review*, 28:3. pp.376-417.
- Dunleavy, P. (1986). Explaining the privatization boom: public choice versus radical approaches. *Public Administration*. 64:1. pp.13-34.
- Dunleavy, P. (1991) *Democracy, Bureaucracy and Public Choice*. Hemel Hempstead: Harvester Wheatsheaf.
- Dunn, W.N. and Holzner, B. (1988) Knowledge in society: anatomy of an emergent field. *Knowledge in Society: The International Journal of Knowledge Transfer*, 1:1. pp. 1-26.
- Dutch Algemene Rekenkamer (1997) *Informatievoorziening en Kengetallen (Information Supply and Indicators)* Algemene Rekenkamer: Den Haag.
- Dutch Social and Cultural Planning Office (2004). *Public Sector Performance. An international comparison*. Den Haag: SCP.

Ehrenberg, R.H. and Stupak, R.J. (2001). Total Quality Management: Its Relations to Administrative Theory and Organizational Behaviour in the Public Sector. in: Stupak, R. J. and Leitner, P. M. *Handbook of public quality management*. New York: Dekker. pp. 40-44.

Eisenstadt, N. (2000). Sure Start: Research into Practice; Practice into Research. *Public Money and Management*. 20:4. pp.6-8.

England, R. E. and Parle, W. M. (1987). Nonmanagerial performance appraisal in large American cities. *Public Administration Review*. 47:6. 498-504.

Epstein, D. (1992). 'Measuring the performance of public services.' in: Holzer (ed.) *Public productivity handbook*. Marcel Dekker: New York. Pp.161-195

Etzioni, E. and Lehman, A. (1967) Some Dangers in 'Valid' Social Measurement: *The Annals of the American Academy of Political and Social Science*, Vol. 373. pp. 1-15.

European Commission (2000) *Communication from the Commission - 27.09.2000: Structural Indicators*. Brussels: European Commission.

European Commission (2000). *The Lisbon European Council - An agenda of economic and social renewal for Europe: Contribution of the European Commission to the special European Council in Lisbon, 23-24th March 2000*. Brussels: European Commission.

European Commission (2001a) *European Governance: a white paper*. Brussels: European Commission.

European Commission (2001b) *Communication from the Commission - 30.10.2001: Structural Indicators*. Brussels: European Commission.

European Commission (2002) *Communication from the Commission - 16.10.2002: Structural Indicators*. Brussels: European Commission.

European Commission (2003) *Communication from the Commission - 08.10.2003: Structural Indicators*. Brussels: European Commission.

European Commission (2004) *Facing the Challenge: The Lisbon Strategy for growth and employment. Report from the High Level Group chaired by Wim Kok*. Brussels: European Commission.

European Commission (2005a) *Delivering on Growth and Jobs: a New and Integrated Economic and Employment Co-ordination cycle in the EU*. Brussels: European Commission.

European Commission (2005b). *Structural Indicators: Update of the Statistical Annex (annex 1) to the 2005 report from the Commission to the spring European Council*. Brussels: European Commission.

- Farr, W. (1852). Influence of Elevation on the Fatality of Cholera. *Journal of Statistical Society of London*. 15:2. pp.155-183.
- Feigenbaum, A.V. (1956), Total quality control, *Harvard Business Review*, 34:6, pp.93-101.
- Ferlie, E., Pettigrew, A., Ashburner, L. and Fitzgerald, L. (1996). *The New Public Management in Action*. Oxford: Oxford University Press.
- Fink, A. (2003). *The Survey Handbook (2<sup>nd</sup> edition)*. Thousand Oaks: Sage.
- Fischer, F. (1990) *Technocracy and the Politics of Expertise*. Newbury Park, CA: Sage.
- Fischer, F. (1993). Citizen Participation and the Democratization of Policy Expertise: From Theoretical Inquiry to Practical Cases. *Policy Sciences*. 26:3. 165-187.
- FitzPatrick P.J. (1960). Leading British Statisticians of the Nineteenth Century. *Journal of the American Statistical Association*. 55:289. pp.38-70.
- Flynn, N. (2002). 'Explaining the New Public Management: The Importance of Context' in McLaughlin, K., Osborne, S. P. and Ferlie, E. (eds) *The New Public Management: Current Trends and Future Prospects*. London: Routledge.
- Flynn, N. and Strehl, F. (1996). *Public Management in Europe*, Hemel Hempstead: Prentice-Hall Harvester Wheatsheaf.
- Fontana, A. and Frey, J. H. (1994). Interviewing: The art of science. In Denzin, N. K. & Lincoln, Y. S. (Eds.), *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Forrester, D.A.R. (1998). *An Invitation to Accounting History*, Glasgow: Strathclyde Convergencies Publishers.
- Frederickson, G. H. (2001) Getting Ranked. *Change*. 33:1. pp.49–55.
- Game, C. (2005). *Comprehensive Performance Assessment in English Local Government: Has life on Animal Farm really improved under Napoleon?* Paper presented at the Conference of the European Group of Public Administration (EGPA) Aug. 31st – Sep. 3rd, 2005, Bern, Switzerland
- Garfield, E. and Welljams-Dorof, A. (1992) Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. *Science & Public Policy*. 19:5. pp.321-327
- Garner, S.P. (1968). Highlights in the development of cost-accounting. in: Chatfield, M. (ed) *Contemporary studies in the evolution of accounting thought*. Belmont (Calif.):Dickenson. pp. 210-222
- GASB (1998) *Performance Measurement at the state and local levels: a summary of survey results*. Washington,DC: GASB.

Geneva Centre for the Democratic Control of Armed Forces and Bulgarian Ministry of Defence. (2001). Workshop on Transparency in Defence Policy, Military budgeting and Procurement. 17-20 May 2001, Sofia.

Giddens, A. (1986) *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.

Glaser, M. (1991). Tailoring performance measurement to fit the organization: from generic to germane. *Public Productivity and Management Review*. 14:3. pp.303-319.

Glass, D.V. (1973) Numbering the people: the eighteenth-century population controversy and the development of census and vital statistics in Britain. Farnborough: Saxon House.

Glass, N. (1999) Sure Start: The Development of an Early Intervention Programme for Young Children in the United Kingdom. *Children and Society*. 13. pp.257-264.

Goddard M. ; Mannion R. and Smith PC. (1999) Assessing the performance of NHS hospital trusts: the role of 'hard'and 'soft'information. *Health Policy*.48:2. pp.119-134.

Goddard, M and Mannion, R. (2004) The role of Horizontal and Vertical approaches to performance measurement and improvement in the public sector. *Public Performance and Management Review*. 28:1. 75-95.

Goldman, A., (2001) Social Epistemology. in: Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy (spring 2001 edition)*, <http://plato.stanford.edu/archives/spr2001/entries/epistemology-social/>>

Goodnow, F.J. (1900:2003). *Politics and administration: a study in government*. New Brunswick (N.J.): Transaction.

Goodsell, C. T. (2004) *The case for bureaucracy: a public administration polemic*. Washington (D.C.): CQ press

Gormley, W.T. and Weimer, D. L.(1999). *Organizational report cards*. Cambridge (Mass.): Harvard university press.

Gould, S.J. (1981). The Politics of the Census. *Natural History*. 90:1 pp.20-23.

Government Performance Project (2003). *Paths to Performance in State and Local Government. A Final Assessment from the Maxwell School of Public Affairs*. Syracuse: Maxwell School of Public Affairs.

- Greener, I. (2003) Performance measurement in the National Health Service: the insistence of measurement and Confusion of Content. *Public Performance and Management Review*. 26:3. pp.237-250
- Grizzle, G. (1987). Linking Performance to Funding Decisions: what is the budgeter's role? *Public Productivity Review*. 10:3. pp.33-44.
- Grizzle, G. (2002). Performance measurement and dysfunction: the dark side of quantifying work. *Public Performance and Management Review*. 25:4. pp.363-369.
- Grizzle, G. and Pettijohn, C. D (2002). Implementing performance based program budgeting: a system dynamic perspective. *Public Administration Review*. 62:1. pp.51-63.
- Gross B.M. (1966). The state of the nation: social systems accounting. in Bauer, R. (ed.) *Social Indicators*. Cambridge, MA: M.I.T. Press.
- Hackman, J. and Oldham, G. (1980). *Work Redesign*. Reading, MA: Addison-Wesley.
- Halachmi, A. (2002). Performance measurement: a look at some possible dysfunctions. *Work study*. 51:5. pp.230 - 239
- Hannan, M. T. and J. Freeman, 1977, 'The Population Ecology of Organizations', *American Journal of Sociology* 82:1. pp.929–964.
- Harari, O. (2001). Ten Reasons Total Quality Management Doesn't Work. in: Stupak, R. J. and Leitner, P. M. *Handbook of public quality management*. New York (N.Y.): Dekker. pp.743-751.
- Hatry, H. (1999) *Performance Measurement: Getting Results*. Washington, D.C: Urban Institute Press.
- Hatry, H. (2002) Performance Measurement: Fashions and Fallacies, *Public Performance and Management Review* 25:4. pp.352–358
- Hay, C. (2002) *Political Analysis: a critical introduction*. Palgrave: Basingstoke.
- Hecht, J. (1977). L'idée de dénombrement jusqu'à la Révolution. in: Bédarida, F., Bouvier, J., Caron, F. Affichard, J. and INSEE (France). *Pour une histoire de la statistique*. Paris: INSEE. pp. 21-83.
- Heckman, J., Heinrich, C., and Smith, J. (1997). Assessing the performance of performance standards in public bureaucracies. *American Economic Review*, 87:2. pp.389-395.
- Hedley, T. P. (1998). Measuring Public Sector Effectiveness Using Private Sector Methods. *Public Productivity and Management Review*, 21:3. pp.251-258.
- Hedström, P. and Swedberg, R. (1998) Social mechanisms : an analytical approach to social theory. *Studies in rationality and social change*. Cambridge: Cambridge University press.

- Heinrich, C.J. (1999) Do Government Bureaucrats Make Effective Use of Performance Management Information? *Journal of Public Administration Research and Theory*. 9:3 363-393.
- Heinrich, C.J. and Lynn, L.E. Jr. (eds) (2000). *Governance and Performance: new perspectives*. Washington, D.C.: Georgetown University Press.
- Helper, H.R. (1857). *The Impending Crisis of the South: How to Meet It*. New York: Burdick Brothers. Electronic Edition of The University of North Carolina at Chapel Hill. <http://docsouth.unc.edu/nc/helper/helper.html>
- Hendricks (1994) Making a splash: Reporting evaluation results effectively. Wholey, J.S. Hatry, H.P. and Newcomer, K.E. (Eds.). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.
- Hendriks, F. and Tops, P. (1999) Between democracy and efficiency: Trends in local government reform in the Netherlands and Germany. *Public Administration* 77:1. pp.133-154.
- Hendriks, F. and Tops, P. (2003). Local public management reforms in the Netherlands: fads, fashions and winds of change. *Public Administration*, 81:2. pp.301-324.
- Hennock, E.P. (1987) The measurement of urban poverty: from the metropolis to the nation, 1880-1920. *Economic History Review*. 40:2 pp.208-227
- Hennock, E.P. (1991). Concepts of poverty in the British social surveys from Charles Booth to Arthur Bowley. in: Bulmer, M., Bales K., and Sklar K.K (eds.) *The Social survey in historical perspective, 1880-1940*, Cambridge: Cambridge University Press. pp. 189-217.
- Henry, G.T. and Dickey, K.C. (2004) Implementing Performance Monitoring: a research and development approach. *Public Administration Review*. 53:3. pp.203-212
- Hill, G.C. (2005) The Effects of Managerial Succession on Organizational Performance. *Journal of Public Administration Research and Theory* 15:4. pp.585-597.
- Hill, H. and Wegener, A. (2002). *Beyond Statistics: Using Quality of Life Indicators for Strategic Management in Local Government*. Gütersloh: Bertelsmann Foundation.
- Hirschman, A. O. (1970) Exit, voice, and loyalty : responses to decline in firms, organizations, and states. Cambridge: Harvard university.
- Ho, A. (2003). Perceptions of Performance Measurement and the Practice of Performance Reporting by Small Cities. *State and Local Government Review*. 35:2. pp.161-173.
- Ho, A. and Coates, P. (2004) Citizen-initiated performance assessment. The initial Iowa Experience. *Public Performance and Management Review*. 27:3. 29-50

Hofstede, G. (1981) Management Control of Public and Not-for-profit activities. *Accounting, Organizations and Society*. 63. pp.193-211.

Hogwood, B.W. (1978) *The primacy of politics in the economic policy of Scottish government* Glasgow: Centre for the Study of Public Policy, University of Strathclyde

Höjer, H. (2001). *Svenska siffror. Nationell integration och identifikation genom statistik 1800-1870 (Sweden by numbers. National integration and identification through statistics, c. 1800–1870)* Uppsala: Uppsala University pp. 261-266

Holsti, O.R (1969) *Content Analysis for the Social Sciences and Humanities*. Reading, Mass: Addison Wesley.

Hoock, J. (1977). D'Aristote à Adam Smith : quelques étapes de la statistique allemande entre le XVIIe et le XIXe siècle. in: Bédarida, F., Bouvier, J., Caron, F. Affichard, J. and INSEE (France). *Pour une histoire de la statistique*. Paris: INSEE. pp.477-493

Hood and Dunsire (1981) *Bureaumerics*. London: Gower.

Hood, C. (1974). Administrative Diseases: Some Types of Dysfunctionality in Administration. *Public Administration*. 52:4. pp.439-454

Hood, C. (1991). A Public Management for All Seasons. *Public Administration Review*. 69:1 pp.3-19.

Hood, C. (2000). *The art of the state: culture, rhetoric, and public management*. Oxford: Clarendon

Hyndman, N. S. and Anderson, R. (1995). The Use of performance information in external reporting; an empirical study of UK executive agencies. *Financial Accountability and Management*. 11:1. pp.1-17.

Ingraham, P.W. (1993). Of Pigs in Pokes and Policy Diffusion: another look at pay for performance. *Public Administration Review*. 53:4. pp.348-356.

Ingraham, P.W. (2005) Performance: Promises to Keep and Miles to Go. *Public Administration Review*. 65:4, pp390-395.

Ingraham, P.W., Joyce, P.G. and Donahue, A.K. (2003) *Government performance : why management matters*. Baltimore (Md.): Johns Hopkins university press.

Innes De Neuville, J.E. (1975). *Social Indicators and Public Policy: Interactive Processes of Design and Application*. Amsterdam: Elsevier.

Innes, J. E. and Booher, D.E. (2000). Indicators for Sustainable Communities: A Strategy Building on Complexity Theory and Distributed Intelligence. *Planning Theory & Practice*. 1:2, pp173-187.



- Innes, J.E. (1990). *Knowledge and Public Policy: The Search for Meaningful Indicators*. New Brunswick: Transaction Publishers.
- Innes, J.E. and Booher, D.E. (2004). Reframing Public Participation: Strategies for the 21st Century. *Planning Theory & Practice*, 5:4. pp.419–436.
- Janis, I.L. (1989) *Crucial Decisions: Leadership in Policymaking and Crisis Management*. New York: Free Press.
- Jardine, L. (2002). *Britain and the Rise of Science*. BBCi History. [www.bbc.co.uk/history](http://www.bbc.co.uk/history) accessed September 2004.
- Jary, D. and Jary, J. (1999). *Unwin Hyman dictionary of sociology*. Glasgow: HarperCollins.
- Johnsen, A. (1999). Implementation mode and local government performance measurement: a Norwegian experience. *Financial Accountability and Management*. 15:1. pp.41-66.
- Johnson, H. T. and Kaplan, R. S. (1987). *Relevance lost: the rise and fall management accounting*. Boston, Mass.: Harvard business school press.
- Jreisat, J. E. (2005) Comparative Public Administration Is Back In, Prudently. *Public Administration Review*, 65:2, pp.231-243.
- Juran, J.M. and Gryna, F.M. (eds) (1988), *Quality Control Handbook, (4th edition)*, McGraw-Hill, New York.
- Kamensky, J. and Morales, A. (2005). *Managing for Results*. Washington DC: IBM center for the business of government.
- Kaplan, R. S. and Norton, D. P. (2001) *The strategy-focused organization : how balanced scorecard companies thrive in the new business environment* Boston (Mass.): Harvard business school press.
- Katz, D. and Kahn, R. L (1966). *The social psychology of organizations*. New York (N.Y.): Wiley.
- Kearney R.C., Feldman B.M. and Scavo C.P.F. (2000). Reinventing Government: City Manager Attitudes and Actions. *Public Administration Review*, 60:6, pp. 535-548
- Kelly, J. M. and Rivenbark, W. C. (2003). *Performance budgeting for state and local government*. Armonk N.Y.: Sharpe.
- Kendall, M.G. (1960). Studies in the History of Probability and Statistics. Where Shall the History of Statistics Begin?. *Biometrika*. 47:3/4. pp.447-449.
- Kettl D.F. (1997). The global revolution in public management: driving themes, missing links. *Journal of Policy Analysis and Management*. 16:4 46-62.

- Kettl D.F. (2000). The Transformation of Governance: Globalization, Devolution, and the Role of Government. *Public Administration Review*, 60:6. pp. 488-497
- Kickert, W. (1997). Public Governance in the Netherlands – an Alternative to Anglo-American “managerialism”. *Public Administration*, 75:4. pp.731-752.
- Kickert, W., (2003). Beyond public management. *Public Management Review*, 5:3. pp.377-400.
- Kimberly J.R. (1976) Organizational size and the structuralist perspective: a review, critique and proposal. *Administrative Science Quarterly*. 21:4. pp.571-597
- Kloby, K. and Kim, Y. (2004) Public Performance Measurement; recent publications. *Public Performance and Management Review*. 28:2 pp.281-289.
- Kopel, S. (2001). Tableaux de bord de mairie; vers l'apprentissage de la responsabilité? (town hall performance indicators; towards learning of responsibility) *Politiques et Management Public*. 19:3. 159-176.
- Kravchuk, R. S. and Schack Ronald W. (1996). Designing effective performance measurement systems under the government performance and results act of 1993. *Public Administration Review*. 56:4. pp.348-358.
- Krippendorff, K. (2004) *Content Analysis: An Introduction to Its Methodology*. 2nd edition, Thousand Oaks, CA: Sage
- Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*. Newberry Park: Sage.
- Kumar, R. (1999) *Research methodology*. London: Sage.
- Laegreid, P. Roness P. G and Rubecksen K. (2005) *Performance management in practice - The Norwegian way*. Paper presented at the Conference of the European Group of Public Administration (EGPA) Aug. 31st – Sep. 3rd, 2005, Bern, Switzerland
- Lawton, A., McKeivitt, D. and Millar, M. (2000). Coping with ambiguity: reconciling external legitimacy and organisational implementation in performance measurement. *Public Money and Management*. 16:3. pp.13-19.
- Lazarsfeld, P.F. (1961). Notes on the History of Quantification in Sociology -- Trends, Sources and Problems. *Isis*. 52:2. pp.277-333.
- Lecuyer, B. (1977). Médecins et observateurs sociaux : les annales d'hygiène publique et de médecine légale (1820-1850). in: Bédarida, F., Bouvier, J., Caron, F. Affichard, J. and INSEE (France). *Pour une histoire de la statistique*. Paris: INSEE. pp. 445-477.

- Lee, R. D. and Burns, R. C. (2000). Performance Measurement in State Budgeting: Advancement and Backsliding from 1990-1995. *Public Budgeting and Finance*. 20:1. pp.38-54.
- Lee, R.M. (2000) *Unobtrusive methods in social research*. Buckingham: pen university press:.
- Lepper, M. R. and Greene, D. (eds.) (1978). *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*. New York; Hillsdale
- Lewis, D.J. and Weigert, A. (1985) Trust as a Social Reality. *Social Forces* 63:4. 1985.
- Light, P.C. (1997). *The tides of reform: making government work, 1945-1995*. New Haven, Conn.: Yale university.
- Lindblom C.E. and Cohen D.K. (1979) *Usable Knowledge: Social Science and Social Problem Solving* Yale University Press
- Lindblom, C.E. (1959). The 'Science' of Muddling Through. *Public Administration Review* 19:2 pp.79-88
- Lindkvist, L. (1996). Performance Based Compensation in health care - a Swedish example. *Financial Accountability and Management*. 12:2. 89-105.
- Lindquist, E.A. (1990) The Third Community, Policy Inquiry, and Social Scientists. In: Brooks, S. and Gagnon A.G. *Social Scientists, Policy, and the State*. New York: Praeger, 21-51.
- Linsley, C.A. and Linsley, C.L. (1993). Booth, Rowntree, and Llewelyn Smith: a reassessment of interwar poverty. *Economic History Review*. 46:1. pp.88-104.
- Lipsky M. (1980) *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- Loeffler, E. and Vintar, M. (eds.) (2004). *Improving the Quality of East and West European Public Services*. Aldershot: Ashgate.
- Lottin, J. (1912) *Quetelet: statisticien et sociologue*. Louvain: Institut supérieur de philosophie.
- Luhmann, N. (1998). Familiarity, Confidence, Trust: Problems and Alternatives. in Gambetta, D. (ed.). *Trust: Making and Breaking Co-operative Relations*. Oxford: Basil Blackwell.
- Machlup (1972) *The production and distribution of knowledge in the United States*. Princeton (N.J.): Princeton university press,
- Mandelbrot, B. B. (1967) How Long is the Coast of Great Britain, Statistical Self Similarity and Fractional Dimension. *Science*, 155. pp.636-638.

- Mani, B.G. (2001). Old Wine in New Bottles Tastes Better: A Case Study of TQM Implementation in the IRS. in: Stupak, R. J. and Leitner, P. M. *Handbook of public quality management*. New York (N.Y.): Dekker. pp. 658-675.
- Manning, P., and Cullum-Swan, B. (1994). Narrative, content, and Semiotic analysis. In Denzin N. and Lincoln Y. *The handbook of qualitative research*. Thousand Oaks, CA: SAGE.
- March, J.G. and Olsen, J.P. (1976) *Ambiguity and Choice in Organizations*. Bergen: Universitetsforlaget.
- March, J.G. and Olsen, J.P. (2004) *The logic of appropriateness*. ARENA Working papers. Oslo: Centre for European Studies.
- Marmot, M. G. (2004). Evidence Based Policy or Policy Based Evidence? Willingness to take action influences the view of the evidence – look at alcohol. *British Medical Journal*. 328. pp. 906-907.
- Marshall M.N., Shekelle P.G., Davies H.T.O., Smith P.C. (2003) Public Reporting on Quality: Lessons from the United States and the United Kingdom. *Health Affairs*. 22:3. pp.134–148.
- Marshall M.N., Shekelle P.G., Leatherman S., Brook R.H. (2000). What do we expect to gain from the public release of performance data? A review of the evidence. *Journal of the American Medical Association*. 283. pp1866–1874.
- Mausolf, C. (2004) Learning from feedback in performance measurement systems. *Public Performance and Management Review*. 28:1. pp.9-29
- Mayo, E. (1933). *The Human Problems of Industrial Civilization*. New York: Macmillan.
- Mayston, David (1985). Non-profit performance indicators in the public sector. *Financial Accountability and Management*. 1:1. pp.51-74.
- McCaffery, J.L. and Jones, L.R. (2004). *Budgeting and Financial Management for National Defence*. Greenwich, Connecticut: Information Age Publishing.
- McDavid, J. C. and Hawthorn, L R. L. (2006) *Program evaluation and performance measurement: an introduction to practice*. Thousand Oaks (Calif.): Sage.
- McGregor, Douglas (1960) *The human side of enterprise* New York (N.Y.): MacGraw-Hill
- McKevitt and Lawton (1996). The Manager, the Citizen, the Politician and Performance Measures. *Public Money and Management*. 16:3. pp.49-54.
- McRae, D.N. (1985). *Policy Indicators*. Chapel Hill: The University of North Carolina Press.

- Meier, K. J., and O'Toole, Jr. L.J. (2002). Public Management and Organizational Performance: The Effect of Managerial Quality. *Journal of Policy Analysis and Management* 21:4 pp.629-643.
- Melkers, J. and Willoughby, K. (1998). The State of the States: performance based budgeting requirements in 47 out of 50. *Public Administration Review*. 58:1. pp.66-73.
- Melkers, J. and Willoughby, K. (2001). Budgeters' views of state performance budgeting systems: distinctions across branches. *Public Administration Review*. 61:1. pp.54-64.
- Merton, R.K. (1968) *Social Theory and Social Structure, expanded edition*. Glencoe, IL: Free Press.
- Meyer, M.W., and Gupta, V. (1994). The performance paradox. *Research in Organizational Behavior*, 16. Pp.309-369.
- Miles, M. and Huberman, M.A. (1994) *Qualitative Data Analysis. 2nd edition*. Thousand Oaks, CA: Sage.
- Miller R.L. and Brewer, J.D. (2003) *The A-Z of social research*. London: Sage.
- Minogue, M., Polidano, C. and Hulme, D. (1998). *Beyond the New Public Management: Changing Ideas and Practices in Governance*, Cheltenham: Edward Elgar.
- Mintzberg, H. (1993). The Pitfalls of Strategic Planning. *California Management Review*, pp. 32-47.
- Mintzberg, H. (1994) *The Rise and Fall of Strategic Planning*. New York: Free Press.
- Modell, S. (2004). Performance Measurement Myths in the Public Sector: a research note. *Financial Accountability and Management*. 20:1. 39-55.
- Moe, Ronald C. (2003). *Administrative Renewal*. Lanham: University Press of America.
- Moesen W. and Persoon A. (2002) Measuring and explaining the productive efficiency of tax offices: A non parametric best practice frontier approach, *Tijdschrift voor Economie en Management*, 47:3. pp.399 - 416.
- Moesen W., (1992) Program evaluation, efficiency measurement and the best practice frontier, in Mayne, J., Bemelmans-Videc, M.J., Hudson, J. and Conner, R. *Advancing public policy evaluation* Amsterdam: North Holland, pp.229-240.
- Moesen, W. and Van Rompuy, V. (1997) *Handboek openbare financiën (Handbook of Public Finances)*. Leuven : Acco.
- Mol, N. (1988) *De allocatiefunctie van de overheidsbegroting*. Twente: Universiteit Twente.

- Mol, N. (1996). Performance indicators in the Dutch department of defence. *Financial Accountability and Management*. 12:1. 71-81
- Moore, M.H. (1998). *Creating Public Value: Strategic Management in Government*. Cambridge, Mas.: Harvard University Press.
- Mosher, F. C. (Ed.). (1975). *American public administration: Past, present, future*. Tuscaloosa: University of Alabama Press
- Mosher, F.C. (1968). *Democracy and the public service*. New York: Oxford University Press.
- Moynihan, D.P. and Ingraham, P.W. (2003). Look for the Silver Lining: Performance Based Accountability Systems Work. *Journal of Public Administration Research and Theory*. 13:4. pp.469-490
- Musgrave, R.A. (1959) *The Theory of Public Finance*. New York: McGraw-Hill.
- Naschold, F. (1996). *New Frontiers in Public Sector Management: Trends and Issues in State and Local Government in Europe*. Berlin: Walter De Gruyter.
- Nathan, R.P. (2004). "Complexifying" Performance Oversight in America's Governments. Association for Public Policy and Management. Presidential Address. Atlanta, G.A, October 29 2004
- NCCD (1996) *A Review of the Literature on Dissemination and Knowledge Utilization*. National Center for the Dissemination of Disability Research: Austin Texas.
- Neuman W.L (1997) *Social Research Methods: Qualitative and Quantitative Approaches, 3rd edition*, Boston: Allyn and Bacon,
- Newcomer, K.E. (1997). Using performance measurement to improve public and nonprofit programs. in Newcomer, K.E. (Ed.). *New Directions for Evaluation*. San Francisco: Jossey-Bass.
- Nichols, P. (1995) *Social survey methods: a fieldguide for development workers. Development guidelines*. Oxford: Oxfam.
- Niskanen, W. (1971) *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.
- Noll, H-H. (1996). *Social indicators and social reporting: the international experience*. Ottawa: social indicators symposium, October 4th and 5th, 1996, Canadian Council on Social Development.
- Noll, H-H. and Zapf, W. (1994) *Social Indicators Research: Societal Monitoring and Social Reporting*. In: Borg, I. and Mohler, P.P. (eds.) *Trends and Perspectives in Empirical Social Research*. Berlin: De Gruyter

- Novick, D. (1973) *Current practice in program budgeting (PPBS): analysis and case studies covering government and business*. London: Heinemann.
- Nyhan, R.C. and Martin, L.L. (1999). Comparative Performance Measurement: a Primer on Data Envelopment Analysis. *Public Productivity and Management Review*. 22:3. pp.348-64.
- OECD (1973). *The OECD Social Indicator Program: List of Social Concerns Common to Most OECD Countries*. Paris: OECD.
- OECD (1976). *Measuring Social Well-being: A Progress Report on the Development of Social Indicators*. Paris: OECD.
- OECD (1982) *The OECD List of Social Indicators*. OECD Social Indicator Development Programme. Paris: OECD.
- OECD (1982) *The OECD List of Social Indicators*. OECD Social Indicator Development Programme. Paris: OECD.
- OECD (1995). *Governance in transition: public management reforms in OECD countries*. Paris: Organization for Economic Co-operation and Development.
- OECD (1996). *Public Management Service. Performance auditing and the modernisation of government*. Paris: Organization for Economic Co-operation and Development.
- OECD (2003) *OECD/World Bank Budget Practices and Procedures Database*. Paris: OECD.
- Osborne, D. and Gaebler, T. (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Reading, Mass.: Addison- Wesley.
- O'Toole, D.E. and Stipak, B. (1988) Budgeting and Productivity Revisited: the Local Government Picture. *Public Productivity Review*. 12:3. pp.1-12.
- O'Toole, L.Jr. (1989). Goal multiplicity in the implementation setting: Subtle impacts and the case of waste-water privatization. *Policy Studies Journal*. 18:1. pp.1-20.
- Paisley, W. (1993). Knowledge utilization: The role of new communications technologies. *Journal of the American Society for Information Science*. pp. 222-234.
- Pallez, F. (2000). De la mesure dans un service public regalien. Peut-on et faut il quantifier la charge de travail des magistrats? (On measurement in a public regulatory service: can we and should we quantify the magistrate's work?) *Politiques Et Management Public*. 18:4. pp.91-118.
- Parsons, W. (1995). *Public policy: an introduction to the theory and practice of policy analysis*. Aldershot : Edwar Elgar.

- Perrin, B. (1998) Effective Use and Misuse of Performance Measurement. *American Journal of Evaluation*. 19:3. pp.367-380.
- Perrot, M. (1977). Premières mesures des faits sociaux: les débuts de la statistique criminelle en France (1780-1830). in: Bédarida, F., Bouvier, J., Caron, F. Affichard, J. and INSEE (France). *Pour une histoire de la statistique*. Paris: INSEE. pp. 233-255.
- Peters, B.G. (1998). *Bringing the State Back in – again*. Ottawa: CCMD.
- Petersen, W. (2000). Counting Heads (Review). *Public Interest*. 141. pp.116-121.
- Platt, J. (2001). Anglo-American contacts in the Development of Research Methods before 1945. in: Bulmer, M., Bales K., and Sklar K.K (eds) (1991). *The Social survey in historical perspective, 1880-1940*. Cambridge: Cambridge University Press.
- Poggie, G. (1965), A main Theme of Contemporary Sociological Analysis; Its Achievements and Limitations, *British Journal of Sociology*. 16:4. pp. 283-294
- Poister, T. H. and Streib, G. (1999). Performance Measurement in Municipal Government: assessing the state of the practice. *Public Administration Review*, 59:4. pp.325-335.
- Poister, T.H. (2003) *Measuring performance in public and nonprofit organizations*. San Francisco (Calif.): Jossey-Bass.
- Poister, T.H. and Streib, G. (1989). Management tools in Municipal Government: Trends over the last decade. *Public Administration Review*, 49:3. pp.240-248.
- Pollit, C. (2000). Is the emperor in is underwear? An analysis of the impacts of public management reform. *Public Management Review*. 2:2. pp.181-199.
- Pollit, C. and Bouckaert, G. (2004) *Public Management Reform. A Comparative Analysis*. Oxford: Oxford University Press.
- Pollitt, C. (1986) Beyond the Managerial Model: The Case for Broadening Performance Assessment in Government and the Public Services. *Financial Accountability and Management*. 2:3. pp.155-170.
- Pollitt, C. (2000). Institutional Amnesia: A Paradox of the 'Information Age'? *Prometheus*. 18:1. pp.5-17.
- Pollitt, C. (2001). Convergence: The Useful Myth? *Public Administration*, 79:4. pp.933-947.
- Ponsaers P. and Bruggeman, W. (2005) De politionele statistische chaos voorbij? (Police Statistics: Beyond the Chaos) *Panopticon*. 2005:2. pp.11-26
- Power, M (1997) *The audit society : rituals of verification*. Oxford: Oxford University Press.



Previts, G.J. and Merino, B.D. (1979). *A history of accounting in America: an historical interpretation of the cultural significance of accounting*. New York: Ronald press.

Raaum, R.B. and Morgan, S.L. (2001). *Performance auditing: a measurement approach*. The institute of internal auditors: Altamonte Springs.

Reck, J.L. (2001). The usefulness of financial and nonfinancial performance indicators in resource allocation decisions. *Journal of Accounting and Public Policy*. 20:1. pp.45-71.

Rejc, A. (2005) Performance measurement in central and eastern Europe, western Europe and North-America: a comparison of research and company practices. UNIDEM Campus Trieste Seminar. 27 June - 1 July 2005.

Rich, R.F. (1997). Measuring knowledge utilization: processes and outcomes, *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization*, 10:3. pp.11–24.

Rich, R.F. and Oh, C.H. (2000) Rationality and Use of Information in Policy Decisions: A Search for Alternatives. *Science Communication*. 22:2. pp.173-211

Richman, J. (2003). Holding Public Health Up for Inspection. in: Costello, J. and Haggart, M. *Public Health and Society*. Basingstoke: Palgrave McMillan.

Ridgeway, V.F. (1956), Dysfunctional Consequences of Performance Measures. *Administrative Science Quarterly* 1:2. pp.240-247.

Rivenbark, W.C. and Pizzarella, C.M. (2002). Auditing performance data in local government. *Public Performance and management review*. 25:4. pp.413-420.

Robbins, D. (1999). Questionnaire Construction. In: Miller, G.J. and Whicker M.L. *Handbook of Research Methods in Public Administration*. New York: Marcel Dekker.

Röber, M. and Löffler, E. (1999). Flexibilities in the German Public Service in: Farnham, D., Horton, S. (Ed.) *Human Resources Flexibilities in the Public Services –International Perspectives*. Macmillan: London.

Rochet, C., Bout-Colonna, L. and Keramidis, O. (2005). *The Risks of Efficiency Indicators in the Monitoring of Public Policies*. Paper presented at the Conference of the European Group of Public Administration (EGPA) Aug. 31st – Sep. 3rd, 2005, Bern, Switzerland.

Rogers, E.M. (1983). *The diffusion of innovations, Second edition*. New York: FreePress.

Romzek, B. and Dubnick, M. (1998) Accountability. In: Shafritz, J.M. *The international encyclopedia of public policy and administration*. Boulder::Westview Press. pp. 6-11.

- Rosenthal Miriam D. (2000). Striving for Perfection: a Brief History of Advances and Undercounts in the U.S. Census. *Government Information Quarterly*. 17:2. pp.193-208.
- Rossi, R. J. and Gilmartin, K.J. (1980). *The handbook of Social Indicators: Sources Characteristics, and Analysis*. New York: Graland STPM Press.
- Roy, C. and Seguin, F. (2000). The institutionalisation of efficiency oriented approaches for public service improvement. *Public Productivity and Management Review*, 23:1. pp.449-468.
- Royal Statistical Society (2003). *Performance indicators: good bad and ugly. Contribution to the OECD World forum on Key Indicators*. Palermo, 10-13 November 2004.
- Rubenstein, R.S., Schwartz, A.E. and Stiefel, L. (2003). Better than Raw: A guide to Measuring Organizational Performance with Adjusted Performance Measures. *Public Administration Review*. 63:5. pp 607–615.
- Rutherford, B.A. (2000). The construction and presentation of performance indicators in executive agency external reports. *Financial Accountability and Management*, 16:3. pp.225-249.
- s.d. (1839). An Account of the Recent Progress and Present Extent of Manufactures in Prussia, and of the Trade of the Prussian Commercial Union in Manufactured Goods *Journal of the Statistical Society of London*. 2:3. pp.135-172.
- Sabatier (1988). An Advocacy Coalition Framework of Policy Change and the role of policy-orientated learning therein *Policy Sciences* 21 pp.129-168.
- Sabatier, P.A. (1986). Top-down and bottom-up approaches to implementation research: Critical analysis and suggested synthesis. *Journal of Public Policy*. 6:1. pp.21-48.
- Sabatier, P.A. (ed.) (1999). *Theories of the Policy Process*. Boulder: Westview.
- Sanderson, I. (2001) Performance management, evaluation and learning in 'modern' local government. *Public Administration*. 79:2. pp.297-313.
- Sanderson, I. (2002). Evaluation, Policy Learning, and Evidence-Based Policy Making. *Public Administration*. 80:1. pp. 1-22.
- Schachter, H.L. (1989) *Frederick Taylor and the public administration community: a reevaluation*. Albany, N.Y.: State university of New York press.
- Schachter, H.L. (2004) Public Productivity in the Classical Age of Public Administration. in: Holzer, M. and Lee S-H.. *Public Productivity Handbook. Second Edition, Revised and Expanded*. New York (NY): Marcel Dekker. pp. 17-31.

- Scheirer, M.A. (1994). Designing and using process evaluation. In Wholey, J.S. Hatry, H.P. and Newcomer, K.E. (Eds.). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass, 40-68.
- Schiavo-Campo, S., and Sundaram, P. (2001) *To serve and to preserve: improving public administration in a competitive world*. Asian Development Bank: Manila.
- Schick, A. (1966). The Road to PPB: The Stages of Budget Reform. *Public Administration Review*. 26:4. pp.243-258.
- Schick, A. (1996) *The Spirit of Reform: Managing the New Zealand State Sectors in a Time of Change*. Report prepared for the State Services Commission. Wellington.
- Schick, A. (1998) Why Most Developing Countries Should Not Try New Zealand Reforms. *The World Bank Research Observer*, 13:1, pp.123–131.
- Schneider A. and Ingram H. (1990) Behavioral assumptions of policy tools. *Journal of Politics*. 52:2. pp.510-529.
- Schneider, M. (2004). Performance Management by Culture in the National Labor Relations Board's Division of Judges and the German Courts of Appeal. *Journal of Public Administration Research and Theory*. 14:1. pp.19-32.
- Schumpeter, J.A. (1972). *History of Economic Analysis*. London: Ruskin House.
- Scott, W.R. (1995). *Institutions and organizations*. Thousand Oaks (Calif.): Sage.
- Senge, P.M. (1990) *The fifth discipline : the art and practice of the learning organization*. Garden City (N.Y.): Doubleday.
- Serban, A.M. and Burke, J.C. (1998). Meeting the performance funding challenge. *Public Productivity and Management Review*, 22:2. pp.157-76.
- Shafritz, J.M., Hyde, A.C. and Parkes, S.J. (2004). *Classics of Public Administration (fifth edition)*. Forth Worth: Harcourt Brace College Publishers.
- Sieber, S. (1981) *Fatal Remedies*. New York: Plenum.
- Silverman, E.B. (2001) *NYPD battles crime : innovative strategies in policing*. Boston (Mass.): Northeastern university press.
- Simon, H.A. (1976) *Administrative behavior : a study of decision-making processes in administrative organization*. New York: Free Press.

- Silverman, D. (1993). *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction*. London: Sage.
- Smith, P. (1988). Assessing competition among local authorities in England and Wales. *Financial Accountability and Management*. 4:3. pp.235-253.
- Smith, P. (1990) The Use of Performance Indicators in the Public Sector. *Journal of the Royal Statistical Society*. 153:1. pp.53-72.
- Smith, P. (1993). Outcome related performance indicators and organizational control in the public sector. *British Journal of Management*, 4:3. pp.135-151.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18. pp.277-310.
- Smith, T.W. (1981). Social Indicators: A Review Essay. *Journal of Social History*. 14:14. pp.739-747.
- Solesbury, W. (2001). *Evidence Based Policy: Whence it Came and Where it's Going*. London: ESRC Centre for Evidence Based Policy and Practice.
- Stehr, N. (1992). *Practical Knowledge: Applying the Social Sciences*. London:Sage.
- Sterck M. and Scheers, B. (2006). *Trends in performance budgeting in seven OECD Countries*. Forthcoming in *Public Performance and Management Review*.
- Sterck, M. and Bouckaert G. (2003). *Internationale trends in Prestatiemeting bij de overheid.(International trends in performance measurement)*. Leuven: SBOV.
- Sterck, M., Scheers, B., and Bouckaert, G. (2005). *The modernization of the public control pyramid: international trends*. Leuven: SBOV.
- Stigler, S.M. (2000) *Statistics on the table: the history of statistical concepts and methods*. Cambridge (Mass.): Harvard university press
- Stiglitz, J.E. (1988) *Economics of the Public Sector (second edition)*. New York; W.W. Norton and Company.
- Stone, A.B. and Stone, D.C. (1975). Early Development of Education in Public Administration. in: Mosher, F.C. (ed.) *American public administration: past, present, future*. University of Alabama press.
- Svara, J. H. (1985). Dichotomy and duality: Reconceptualizing the relationship between policy and administration in council-manager cities. *Public Administration Review*, 45:1. pp.221-32.
- Swiss, J.E. (1992). Adapting Total Quality Management to government. *Public Administration Review*. 52:4. pp.356-362.

Taylor Fitz-Gibbon, C. (2002). Evaluation in the Age of Indicators. Challenges for Public Sector Management. *Evaluation*. 8:1. pp.140-148.

Ter Bogt, H. (2004). Politicians in search of performance information? Survey research on Dutch alderman's use of performance information. *Financial Accountability and Management* 20:3. pp.221-252.

The Working Party on Performance Monitoring in the Public Services of the Royal Statistical Society (2005). *Performance indicators: good bad and ugly*. *Journal of the Royal Statistical Society*. 168:1. pp.1-27.

Thierry, H. and Nederlandse Vereniging voor Bedrijfspsychologie (1977) *Sociale Indicatoren in Beweging*. Deventer: Kluwer.

Thiry, J.P. and Meyntjens, C. (1979). *Ontwikkeling van sociale indicatoren. Naar een sociaal verslag. (Development of Social Indicators; towards a social report)*. Brussel: Planbureau.

Thompson, J.R (2000). Reinvention As Reform: Assessing the National Performance Review. *Public Administration Review*. 60:6. pp.508-522

Thompson, J.R. (2000). The dual potentialities of performance measurement. *Public Productivity and Management Review*, 23:3. pp.267-81.

Thompson. J.R. (2001) The Clinton Reforms and the Administrative Ascendancy of Congress. *The American Review of Public Administration*. 31:3. pp.249-272.

Tucker, C.M. (2003) *The Lisbon Strategy and the Open Method of Coordination: a New Vision and the Revolutionary Potential of Soft Governance in the European Union*. American Political Science Association. 2003 annual meeting.

Tulkens, H. (1993) On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit", *Journal of Productivity Analysis* 4:(1/2). Pp.183-210.

U.K. Cabinet Office. (1999). *Modernising Government*. London: Stationery Office.

U.K. House of Commons, Public Administration Select Committee (2003). *On Target Government by Measurement. Fifth report of Session 2002-2003 (vol1)*. London: Stationery Office.

U.K. Office for National Statistics. (2001). *200 years of Census*. London: ONS.

U.S. Bureau of the Census (1989). *200 Years of Census Taking: Population and Housing Questions, 1790-1990*. Washington, DC: Bureau of the Census.

U.S. Bureau of the Census (2000). *Fact finder for the Nation: History and Organization*. Washington DC: Bureau of the Census.

U.S. General Accounting Office (1997). *Performance Budgeting: Past Initiatives Offer Insights for GPRA Implementation*. Washington DC: GAO.

U.S. General Accounting Office (2000) *Reinventing Government: Status of NPR Recommendations at 10 Federal Agencies*. Washington D.C: GAO.

U.S. National Performance Review (1993). *Report of the National Performance Review, From Red Tape to Results; creating a government that works better and costs less*. Washington DC: US Government Printing Office.

United Nations (1975). *Towards a System of Social and Demographic Statistics*. New York: United Nations.

United Way of America (1996) *Measuring Program Outcomes: A Practical Approach*. Alexandria, VA: United Way of America.

Van de Walle S. (2004) *Perceptions of administrative performance: the key to trust in government?* K.U.Leuven. Faculteit sociale wetenschappen. Departement Politieke Wetenschappen.

Van De Walle, S. (2005) *Measuring Bureaucratic Quality in Governance Indicators* EGPA annual conference. Study Group II: Productivity and Quality in the Public Sector. 31 augustus - 2 september 2005. Bern: Zwitserland.

Van Dooren W. (2005) Prestatiemeting. Een toepassing voor politiekorpsen. In: Ponsaers P., Enhus E., Easton M., Bruggeman W. and Collier A. (ed.). *10 miljoen klanten moet je verdienen*. Brussel: Politea. pp. 29-45.

Van Dooren, W. and Bouckaert, G. (2004). *Prestatiemeting in Vlaanderen: Vragen voor een informatiebeleid. (Performance measurement in Flanders: issues for an information policy)*. Leuven: SBOV.

Van Dooren, W. and Van De Walle, S. (2004) Why do Belgian public agencies use the Common Assessment Framework (CAF)? in: Loeffler, E. and Vintar, M. (eds.) (2004). *Improving the Quality of East and West European Public Services*. Aldershot: Ashgate.

Van Dooren, W. and Van de Walle, S. (2004) Zelfevaluatie of zelfpromotie? Een onderzoek naar de motieven voor het gebruik van een kwaliteitsmodel: het Common Assessment Framework. *Bestuurskunde*. 13:2. pp.87-98.

Van Dooren, W., Bouckaert, G. and Thys, N., (2004) Quality Management and Management of quality in the European Public Administrations. in: Loeffler, E. and Vintar, M. (eds.) (2004). *Improving the Quality of East and West European Public Services*. Aldershot: Ashgate.

Van Dooren, W., Sterck M. and Van de Walle S. (2005) "Ver(on)gelijkt? De mogelijkheden en moeilijkheden van internationale indicatoren In: Bouckaert Geert, Hondeghe Annie, Janvier Ria, De Rynck Filip & Bursens Peter (Red). Jaarboek SBOV 2004: Vlaanderen en/in Europa: omgaan met de Unie en vergelijkend leren. Leuven: SBOV pp. 11-28

Van Reeth, Wouter (2002) *The bearable lightness of budgeting : the uneven implementation of performance oriented budget reform across agencies* Leuven: KUL. Faculteit sociale wetenschappen. Departement politieke wetenschappen

Van Thiel, S. and Leeuw, F. (2002) The Performance Paradox in the Public Sector. *Public Performance & Management Review*. 25:3. pp.267-281

Vanden Eeckhout, P., Tulkens H., Jamar M.A. (1993) Cost Efficiency in Belgian Municipalities, The Measurement of Productive Efficiency: Techniques and Applications. in Fried, H.O., Lovell, C.A.K and Schmidt, S.S. (eds.), *The measurement of productive efficiency: techniques and applications*. Oxford University Press.

Varone, F., Jacob, S., and De Winter, L. (2005). Polity, Politics and Policy Evaluation in Belgium. *Evaluation*. 11:3. pp.253-273.

Verhoest, K. (2002) *Resultaatgericht verzelfstandigen : een analyse vanuit een verruimd principaal-agent perspectief, de invloed van nieuwsoortige sturingsmechanismen op de resultaten van uitvoerende overheidsorganisaties. Results oriented autonomization: an analysis from an expanded principal-agent perspective, the influence of novel steering instruments on the results of executive agencies*. Leuven: KULeuven. Faculteit sociale wetenschappen. Departement politieke wetenschappen.

Waintrop, F. (2004). *France: le pari d'une réforme systémique: les enjeux de la réforme budgétaire*. Conference of the European Group of Public Administration (EGPA) Sep. 1st – Sep. 3rd, 2004. Ljubljana, Slovenia.

Wang, X. (2002). Assessing Performance Measurement Impact; a study of US local governments. *Public Performance and Management Review* , 26:1. pp.26-43.

Wang, X. and Berman, E. (2000). Hypotheses about performance measurement in counties: findings from a survey. *Journal of Public Administration and Theory* , 11:3. pp.403-28.

Webb, E., Campbell, D. Schwartz, R. and Sechrest, L. (1966). *Unobtrusive measures: nonreactive research in the social sciences*. Boston: Houghton Mifflin.

Weiss, C.H, (ed.) (1977) *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.

- Weiss, C.H. (1979). The many meanings of research utilization. *Public Administration Review*. 39:5. pp.426-431.
- Weiss, C.H. (1980). Knowledge creep and decision accretion *Knowledge: Creation, Diffusion and Utilization*. 1. Pp.381-404.
- Weiss, C.H. (1998) *Evaluation (second edition)*. Upper Saddle River: Prentice Hall.
- Weiss, C.H. (1999). The Interface between Evaluation and Public Policy. *Evaluation*. 5:4. pp.468-486.
- Wholey J.S. (1994) Assessing the feasibility and likely usefulness of evaluation, in Wholey J.S., Hatry H., Newcomer K.E. (eds.). *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass.
- Wholey, J. (1999). Performance based management: responding to the challenges. *Public Productivity and Management Review*, 22:3, pp. 288–307.
- Wholey, J. and Hatry, H. (1992) The Case for Performance Monitoring. *Public Administration Review*. 52:6. pp.604-610
- Wholey, J.S. (1997). Trends in performance measurement: Challenges for evaluators. In Chelimsky, E. and Shadish, W.R. (Eds.) *Evaluation for the 21st century: A handbook*. Thousand Oaks, CA: Sage Publications.
- Wholey, J.S. Hatry, H.P. and Newcomer, K.E. (Eds.) (1994). *Handbook of practical program evaluation*. San Francisco: Jossey-Bass.
- Wiggins, A. and Tymms, P. (2002). Dysfunctional Effects of League Tables: a comparison between English and Scottish Primary Schools. *Public Money and Management*. 22:1. Pp.43-48.
- Wilcox, W. (1934). Note on the Chronology of Statistical Societies. *Journal of the American Statistical Association*. 29:188. pp. 418-420.
- Wildavsky, A. (1969). Rescuing Policy Analysis from PPBS. *Public Administration Review* 29:2 pp.189-202.
- Williams, D.W. (2000). Reinventing the Proverbs of Government. *Public Administration Review*. 60:6. pp.522-535.
- Williams, D.W. (2002). Before Performance Measurement. *Administrative Theory and Praxis*. 24:3. pp.457-486.
- Williams, D.W. (2003). Measuring Government in the Early Twentieth Century. *Public Administration Review*. 63:6. pp.643-659.



- Williams, D.W. (2004). Evolution of performance measurement until 1930. *Administration and Society*. 36:2. 131-165.
- Williams, W. (1998). *Honest Numbers and Democracy*. Washington DC: Georgetown University Press.
- Willoughby, W.F. (1918). *The Movement for Budgetary Reform in the States*. New York: D. Appleton and Company for the Institute for Government Research.
- Wilson, D. (2004) Which ranking? The impact of a 'value added' measure of secondary school performance. *Public Money and Management*. 24:1. pp.37-45.
- Wilson, J.Q. (1989). *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.
- Winston, J. (1993) Performance indicators: do they perform? *Evaluation News and Comment*. 2:3, pp. 22-29
- Woodbury, K., Dollery, B. and Rao, P. (2003). Is local government efficiency measurement in Australia adequate? An analysis of the evidence. *Public Performance and Management Review*. 27:2. pp.77-91.
- Worthington, A.C. and Dollery, B.E. (2000). Measuring Efficiency in Local Governments' Planning and Regulatory Function. *Public Productivity and Management Review*. 23:4. pp.469-85.
- Yin, K. (1994) *Case study research: design and methods. Second edition*. Thousand Oaks (Calif.): Sage.
- Young, K., Ashby, D., Boaz, A. and Grayson, L., (2002). Social Sciences and the Evidence-Based Policy Movement. *Social Policy and Society*. 1:3. pp.215-224.
- Zifcak, S (1994). *New managerialism: administrative reform in Whitehall and Canberra*. Open university press: Buckingham.



## 13. Appendices

### 13.1. Appendices to the literature study

#### 13.1.1. List of articles

1. Alcock, P. (2004) Targets, indicators and milestones: what is driving area based policy in England? *Public Management Review*. 6:2. 211-227
2. Ammons, D. N. and Rodriguez, A. (1986). Performance appraisal for upper management in city governments. *Public Administration Review* , 46:5. 460-467.
3. Ammons, D. N., Coe, C. C. and Lombardo, M. (2001). Performance comparison projects in local government: participants' perspectives. *Public Administration Review* , 61, (1): 100-110.
4. Ammons, D.N. (2002). Performance and managerial thinking. *Public Performance and Management Review*. 25:4. 344-347.
5. Ammons, D.N. (1985). Common barriers to productivity Improvement in local government. *Public Productivity Review*. 9. 187-202.
6. Andrews, R. (2004) Analysing deprivation and local authority performance: the implications for CPA. *Public Money and Management*. 24:1. 19-26.
7. Behn, R. D. (2003) Why measure performance? Different purposes require different measures. *Public Administration Review*. 63:5. 586-606.
8. Behn, R. D. (2002). The psychological barriers to performance measurement: Or why isn't everyone jumping on the performance management bandwagon? *Public Performance and Management Review*. 26:1. 5-25.
9. Behn, R. D. And Kant, P. A. (1999). Strategies for avoiding the pitfalls of performance contracting. *Public Productivity and Management Review*. 22:4. 470-489.
10. Berman, E. And Wang, X. (2000). Performance measurement in U.S. Counties: Capacity for reform. *Public Administration Review*. 60:5. 409-20.
11. Berman, E. (2002). How Useful is performance management? *Public Performance and Management Review*. 25:4. 348-351.
12. Berry, F. S., Brower, R. S. and Flowers, G. (2000). Implementing performance accountability in Florida: what changed, what mattered, what resulted? *Public Productivity and Management Review*. 23:3. 338-358.
13. Bouckaert, G. (1993) Measurement and meaningful management. *Public Productivity and Management Review*. 17:1. 31-43.
14. Bouckaert, G. and Peters, B.G (2002). Performance Measurement and Management. The Achilles' Heel in administrative modernization. *Public Performance and Management Review*. 25:4.
15. Bouckaert, G. (1990). The history of the productivity Movement. *Public Productivity and Management Review*. 14:1.

16. Boyne, G. And Law, J. (1991). Accountability and local authority annual reports the case of the Welsh District councils. *Financial Accountability and Management*. 7:3. 179-94.
17. Broom, C. A. (1995). Performance-based government models: building a track-record. *Public Budgeting and Finance*. 15:4. 3-17.
18. Brown, R.E. and Pyers, J.B. (1988). Putting teeth in efficiency and effectiveness of public services. *Public Administration Review*. 48:3. 735-742
19. Carter, N. (1991). Learning to measure performance: the use of indicators in organizations. *Public Administration*. 69:1. 85-101.
20. Chalos, P. and Cherian, J. (1995). An application of Data Envelopment Analysis to public sector performance measurement and accountability. *Journal of Accounting and Public Policy*. 14:2. 143-160
21. Coe, C. K. (2004). A report card on report cards. *Public Performance and Management Review*. 27:2. 53-76.
22. Dawson, D. And Street, A. (2000). Comparing NHS Hospital unit costs. *Public Money and Management*. 20:4. 58-62.
23. de Lancer Julnes, P. (2001). Does participation increase perceptions of usefulness?: An evaluation of a participatory approach to the development of performance measures. *Public Performance and Management Review*. 24:4. 403-418.
24. de Lancer Julnes, P. And Holzer, M. (2001). Promoting utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review*. 61:6. 650-665.
25. England, R. E. and Parle, W. M. (1987). Nonmanagerial performance appraisal in large American cities. *Public Administration Review*. 47:6. 498-504.
26. Glaser, M. (1991). Tailoring performance measurement to fit the organization: from generic to germane. *Public Productivity and Management Review*. 14:3. 303-319.
27. Goddard, M and Mannion, R. (2004) The role of Horizontal and Vertical approaches to performance measurement and improvement in the public sector. *Public Performance and Management Review*. 28:1. 75-95.
28. Greener, Ian (2003) Performance measurement in the National Health Service: the insistence of measurement and Confusion of Content. *Public Performance and Management Review*. 26:3. 237-250
29. Grizzle, G. and Pettijohn, C. D (2002). Implementing performance based program budgeting: a system dynamic perspective. *Public Administration Review*. 62:1. 51-63
30. Grizzle, G (2002). Performance measurement and dysfunction: the dark side of quantifying work. *Public Performance and Management Review*. 25:4. 363-369
31. Grizzle, G. (1987). Linking Performance to Funding Decisions: what is the budgeter's role? *Public Productivity Review*. 10:3. 33-44
32. Halachmi, A. (2002). Performance measurement: a look at some possible dysfunctions. *Work study*. 51:5. 230 - 239

33. Hatry, H. (2002). Performance measurement; fashions and fallacies. *Public Performance and Management Review*. 25:4, 352-358
34. Hedley, T. P. (1998). Measuring Public Sector Effectiveness Using Private Sector Methods. *Public Productivity and Management Review*, 21:3. 251-258.
35. Heinrich, C.J. (1999) Do Government Bureaucrats Make Effective Use of Performance Management Information? *Journal of Public Administration Research and Theory*. 9:3 363-393.
36. Henry, G.T. and Dickey, K.C. (2004) Implementing Performance Monitoring: a research and development approach. *Public Administration Review*. 53:3. 203-212
37. Ho, A. and Coates, P. (2004) Citizen-initiated performance assessment. The initial Iowa Experience. *Public Performance and Management Review*. 27:3. 29-50
38. Ho, A. (2003). Perceptions of Performance Measurement and the Practice of Performance Reporting by Small Cities. *State and Local Government Review*. 35:2. 161-173.
39. Hyndman, N. S. and Anderson, R. (1995). The Use of performance information in external reporting; an empirical study of UK executive agencies. *Financial Accountability and Management*. 11:1. 1-17.
40. Ingraham, P. (1993). Of Pigs in Pokes and Policy Diffusion: another look at pay for performance. *Public Administration Review*. 53:4. 348-356.
41. Johnsen, A. (1999). Implementation mode and local government performance measurement: a Norwegian experience. *Financial Accountability and Management*. 15:1. 41-66.
42. Kravchuk, R. S. and Schack Ronald W. (1996). Designing effective performance measurement systems under the government performance and results act of 1993. *Public Administration Review*. 56:4. 348-58.
43. Lawton, A., McKeivitt, D. and Millar, M. (2000). Coping with ambiguity: reconciling external legitimacy and organisational implementation in performance measurement. *Public Money and Management*. 16:3. 13-19.
44. Lee, R. D. and Burns, R. C. (2000). Performance Measurement in State Budgeting: Advancement and Backsliding from 1990-1995. *Public Budgeting and Finance*. 20:1. 38-54.
45. Lindkvist, L. (1996). Performance Based Compensation in health care - a Swedish example. *Financial Accountability and Management*. 12:2. 89-105.
46. Mausolf, C. (2004) Learning from feedback in performance measurement systems. *Public Performance and Management Review*. 28:1. 9-29
47. Mayston, David (1985). Non-profit performance indicators in the public sector. *Financial Accountability and Management*. 1:1. 51-74.
48. McKeivitt and Lawton (1996). The Manager, the Citizen, the Politician and Performance Measures. *Public Money and Management*. 16:3. 49-54.
49. Melkers, J. and Willoughby, K. (1998). The State of the States: performance based budgeting requirements in 47 out of 50. *Public Administration Review*. 58:1. 66-73.
50. Melkers, J. and Willoughby, K. (2001). Budgeters' views of state performance budgeting systems: distinctions across branches. *Public Administration Review*. 61:1. 54-64.

51. Modell, S. (2004). Performance Measurement Myths in the Public Sector: a research note. *Financial Accountability and Management*. 20:1. 39-55.
52. Mol, N. (1996). Performance indicators in the Dutch department of defence. *Financial Accountability and Management*. 12:1. 71-81
53. Moynihan, D. P. and Ingraham, P. W. (2003). Look for the Silver Lining: Performance Based Accountability Systems Work. *Journal of Public Administration Research and Theory*. 13:4. 469-490
54. Nyhan, R. C. and Martin, L. L. (1999). Comparative Performance Measurement: a Primer on Data Envelopment Analysis. *Public Productivity and Management Review*. 22:3. 348-64.
55. O'Toole, D.E. and Stipak, B. (1988) Budgeting and Productivity Revisited: the Local Government Picture. *Public Productivity Review*. 12:3. 1-12.
56. Poister, T. H. and Streib, G. (1999). Performance Measurement in Municipal Government: assessing the state of the practice. *Public Administration Review*, 59:4. 325-335.
57. Poister, T. H. and Streib, G. (1989). Management tools in Municipal Government: Trends over the last decade. *Public Administration Review*, 49:3. 240-248.
58. Reck, J. L. (2001). The usefulness of financial and nonfinancial performance indicators in resource allocation decisions. *Journal of Accounting and Public Policy*. 20:1. 45-71.
59. Rivenbark, W.C. and Pizzarella, C.M. (2002). Auditing performance data in local government. *Public Performance and management review*. 25:4. 413-420.
60. Roy, C. and Seguin, F. (2000). The institutionalisation of efficiency oriented approaches for public service improvement. *Public Productivity and Management Review*, 23:1. 449-468.
61. Rubenstein, R.S., Schwartz, A. E. and Stiefel, L. (2003). Better than Raw: A guide to Measuring Organizational Performance with Adjusted Performance Measures. *Public Administration Review*. 63:5.
62. Rutherford, B. A. (2000). The construction and presentation of performance indicators in executive agency external reports. *Financial Accountability and Management*, 16:3. 225-249.
63. Sanderson, I. (2001) Performance management, evaluation and learning in 'modern' local government. *Public Administration*. 79:2. 297-313.
64. Schneider, M. (2004). Performance Management by Culture in the National Labor Relations Board's Division of Judges and the German Courts of Appeal. *Journal of Public Administration Research and Theory*. 14:1. 19-32.
65. Serban, A. M. and Burke, J. C. (1998). Meeting the performance funding challenge. *Public Productivity and Management Review*, 22:2. 157-76.
66. Smith, P. (1988). Assessing competition among local authorities in England and Wales. *Financial Accountability and Management*. 4:3. 235-253.
67. Smith, P. (1993). Outcome related performance indicators and organizational control in the public sector. *British Journal of Management*, 4:3. 135-151.
68. Smith, P. (1990) The Use of Performance Indicators in the Public Sector. *Journal of the Royal Statistical society*. 153:1. 53-72.

69. Ter Bogt, H. (2004). Politicians in search of performance information? Survey research on Dutch alderman's use of performance information. *Financial Accountability and Management* 20:3. 221-252.
70. Thompson, J. R. (2000). The dual potentialities of performance measurement. *Public Productivity and Management Review*, 23:3. 267-81.
71. van Thiel, S. and Leeuw, F.L. (2002). The Performance Paradox in the Public Sector. *Public performance and management review*. 25:3. 267-281
72. Wang, X. (2002). Assessing Performance Measurement Impact; a study of US local governments. *Public Performance and Management Review* , 26:1. 26-43.
73. Wang, X. and Berman, E. (2000). Hypotheses about performance measurement in counties: findings from a survey. *Journal of Public Administration and Theory* , 11:3. 403-28.
74. Wholey, J. and Hatry, H. (1992) The Case for Performance Monitoring. *Public Administration Review*. 52:6. 604-610
75. Wiggins, A. and Tymms, P. (2002). Dysfunctional Effects of League Tables: a comparison between English and Scottish Primary Schools. *Public Money and Management*. 43-48.
76. Williams, Daniel W. (2003). Measuring Government Efficiency in the Early Twentieth Century. *Public Administration Review*. 63:6. 643-659.
77. Williams, D. W. (2004). Evolution of performance measurement until 1930. *Administration and Society*. 36:2. 131-165.
78. Wilson, D. (2004) Which ranking? The impact of a 'value added' measure of secondary school performance. *Public Money and Management*. 24:1. p37-45.
79. Woodbury, K., Dollery, B. and Rao, P. (2003). Is local government efficiency measurement in Australia adequate? An analysis of the evidence. *Public Performance and Management Review*. 27:2. 77-91.
80. Worthington, A. C. and Dollery, B. E. (2000). Measuring Efficiency in Local Governments' Planning and Regulatory Function. *Public Productivity and Management Review*. 23:4. 469-85.

### 13.1.2. Analysis of the studies: research questions and conclusions

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Alcock, Peter	2004				What are the effects of target setting in Area Based Initiatives?	<ul style="list-style-type: none"> <li>a. Health Action Zones are subjected to performance management practices.</li> <li>This incurred problems.</li> <li>b. The NHS shifted targets</li> <li>c. Creation of a bureaucracy and duplication</li> <li>d. Problems with quantification</li> <li>e. Tendency towards inputs and outputs, not outcomes</li> <li>f. Gearing activity towards the indicators and the milestone</li> <li>g. Setting soft targets</li> <li>h. Picking winners</li> <li>i. This was behavior was stimulated by the high political profile of the program</li> <li>j. The process as a target (e.g. participation) in itself is neglected</li> </ul>
Ammons and Rodriguez	1986	Does organization size and organization form (council-manager/mayor-council) explain the use of appraisal techniques?	<p>Is the appraisal system formal, informal or non-existent?</p> <p>What are the objectives of the appraisal system?</p> <p>What is the frequency of appraisal?</p>	<p>What appraisal techniques are in use?</p> <p>What is the aggregate amount of time devoted to the evaluation?</p>	<p>What is the satisfaction with performance appraisal for managerial performance?</p>	<p>MBO is the most popular appraisal system. Satisfaction with the appraisal systems is high, though there are some deficiencies:</p> <ul style="list-style-type: none"> <li>a. Only 59% of the evaluations have a formal documented basis.</li> <li>b. 12% has no evaluation at all.</li> <li>c. 16 % of the evaluations has a full reliance on rating scales.</li> <li>d. Only very modest amounts of executive and staff time is devoted to the process.</li> <li>e. Size does not influence the use of techniques; the form does</li> </ul>
Ammons, Coe, and Lombardo	2001			How can performance comparison projects be made more valuable for participators?		<ul style="list-style-type: none"> <li>a. Tune expectations of participants and expected results: benchmarking is a tool, not a promise</li> <li>b. Participants recommended a broader range of services, more attention to cost accounting, more sophisticated measures, to rely more on data as submitted (speed vs. quality) and more in depth analysis as a next step</li> </ul>



Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Ammons, D.N	1985	What are the barriers to productivity improvement programs?				<ul style="list-style-type: none"> <li>a. List of 37 barriers to performance improvement</li> <li>b. These barriers are identified in order to tackle them, not to abdicate</li> </ul>
Ammons, D.N	2002				Why is there such an imbalance between promised and actual results in terms of service improvement?	<ul style="list-style-type: none"> <li>a. Performance measurement is used for two reasons: accountability and service improvement. The first is not much contested. The latter however is highly controversial.</li> <li>b. The reasons for the imbalance between promise and results are: <ul style="list-style-type: none"> <li>- poor design for improvement (simplicity as the highest operational priority)</li> <li>- small gains are unheralded</li> <li>- other factors get the credit for large performance gains</li> </ul> </li> <li>c. Research and practice should take the design into account when assessing performance measurement</li> </ul>
Andrews	2004			Should deprivation be a part of the comprehensive performance assessment in the UK?		<ul style="list-style-type: none"> <li>a. Deprivation has a statistically significant impact on outcomes</li> <li>b. CPA should acknowledge these external constraints</li> </ul>
Behn	2003		What are the implications for the selection of the measures?	What are the main purposes of performance measurement?		<ul style="list-style-type: none"> <li>a. Eight uses are identified; to evaluate, to control, to budget, to promote, to celebrate, to learn, to motivate, and to improve</li> <li>b. The different uses require different measures; there is no one best measure</li> </ul>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Behn	2002	Why is performance management mostly only used in a rhetorical way?				<p>a. There are practical, political, managerial and psychological reasons for the often rhetorical use of performance information.</p> <p>b. The psychological reasons are fear for the consequences and the need for a mental reorientation. This article reports on the latter reason.</p> <p>c. The mental reorientation requires a new way of citizen thinking, legislative thinking, public employee thinking, policy thinking, assistant secretary thinking, distrustful thinking and big picture thinking.</p> <p>d. To overcome these psychological barriers, government needs performance leadership.</p>
Behn and Kant	1999		How to avoid the pitfalls in performance contracting?		What are the pitfalls in performance contracting?	<p>a. The pitfalls in performance contracting, opposed to regulatory contracting consist of three punishment pitfalls (inhibit experimentation, encourage cost cutting rather than innovation, stifle overachievement), three revolving door pitfalls (not provide for start-up cost, inhibit symbiotic relationships, reward promises and not performance) and four complexity pitfalls (reliance on output and not outcome, distort behavior, creaming, undermine equity and fairness).</p> <p>b. Eight strategies that describe how to design and implement performance contracts are put forward.</p>
Berman and Wang	2000		To what extent do counties use some form of performance measurement?	Which capacities must be present for different levels of implementation and success? How can capacity be increased?		<p>a. 33% of the counties use some form performance measurement</p> <p>b. 6,8% can be said to have "high use"</p> <p>c. One third have a high capacity and capacity is correlating positively with use</p> <p>d. Capacity requires that jurisdictions are able to relate outputs to operations, to collect timely data, have capable staff, adequate information systems and support from department heads and elected officials.</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Berman, Evan	2002		How useful is performance measurement?			<p>a. There is a limited use of performance measurement. Success stories are rare.</p> <p>b. Compared to other productivity improvement tools, performance measurement may be said to be halfway its full implementation or failure (ten years out of twenty).</p> <p>c. Two midcourse adjustments are proposed:</p> <ul style="list-style-type: none"> <li>- more reliance on IT</li> <li>- more investment in the capacity to collect outcome data</li> </ul>
Berry, Brower and Flowers	2000	What factors make a difference in performance accountability implementation?	How has the performance accountability initiative changed management and budgeting?		Is the Performance Accountability Initiative creating better outcomes for Government?	<p>a. Management and budgeting has changed in an incremental rather than a frame braking way; line item budgets have been sustained, only minimal incentive packages exist. Nevertheless, PB is being used. The authors suggest that the theoretical implications might be that normative and cognitive institutionalism may explain PB better than agency theory.</p> <p>b. The most important factor for adopting and implementing PB is the explicit connection between PB and strategic management.</p> <p>c. Routine technologies have easier and more coherent implementations of PB accountability</p> <p>d. The congruency of external interest group's agenda with regard to the agency's mission supports PB implementation.</p> <p>e. Effects for the organization are chiefly a stronger focus on effects instead of output, increased organizational learning, and better coordination across agencies.</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Bouckaert	1993			How to design performance measurement systems that are "fit for use" for contemporary management?		<p>a. Performance measurement is a dynamic concept:</p> <ul style="list-style-type: none"> <li>- performance measurement systems are used more extensively (in more services and policy fields),</li> <li>- the evaluation of the public sector has evolved from a more juridical approach towards a performance based approach</li> <li>- performance measurement is increasingly used by the management itself.</li> </ul> <p>b. These changes affect the performance measurement systems: the importance of technical validity is supplemented by legitimacy and functionality.</p> <p>c. The combination of the three dimension allows for distinguishing different profiles of performance measurement systems: from the worst case to a global optimum.</p>
Bouckaert and Peters	2002		What are the pitfalls in performance management?	What are the pitfalls in performance measurement?	What are the possible negative effects of performance measurement?	<ol style="list-style-type: none"> <li>1. Costs appear in the short term and are tangible, benefits appear in the longer term and are not tangible</li> <li>2. Performance measurement at micro level may cause an illusion of control, compared to the macro level</li> <li>3. Quality as a complement of quantity</li> <li>4. What is the relation between performance and resources?</li> <li>5. Not everyone can be a best practice</li> <li>6. Learning across organizations and countries is important</li> <li>7. societal performance is more than the sum of its components</li> <li>8. Tension between performance and accountability</li> <li>9. performance contracts or investing in trust</li> <li>10. The triangle of? trust, performance and satisfaction is not clear</li> <li>11. Pathologies need to be studied</li> </ol>
Bouckaert, Geert	1990	How does the spirit of the age influence the productivity concerns?	What is the history of the productivity movement?	What is the history of the productivity movement? What are the methodological shifts?		<p>a. Productivity concerns in the USA go back at least a centuryb. the spirit of the age determines the concept of productivityc. methodologically, productivity becomes increasingly complex.</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Boyne and Law	1991		To what extent are performance indicators used in local authorities annual reports?			<p>a. The annual reports on performance are generally of poor performance, most of the performance indicators refer to service input while broader issues concerning 'citizenship' and equity of service receive little attention</p> <p>b. There is no improvement in the annual reports over time</p> <p>c. It is concluded that;</p> <ul style="list-style-type: none"> <li>- a new code of practice should be issued by central government,</li> <li>- reports should be brought overtly in the political arena</li> <li>- the greatest weakness is the failure to state explicit targets and priorities.</li> </ul>
Broom	1995		Are PM initiatives being sustained over time?	What are the (design) factors critical to sustained success?		<p>Factors critical to sustainable success:</p> <ol style="list-style-type: none"> <li>a. First things first: identify the need</li> <li>b. Visualize how the products will be used</li> <li>c. The desire to learn must be present</li> <li>d. The use of performance information for budget decision needs not to change the budget process, but needs to inform the deliberations</li> <li>e. The executive and legislative branch need to question agencies about performance</li> <li>f. Performance based government efforts need adequate time</li> <li>g. Executive and legislative leadership is crucial</li> <li>h. Performance based efforts can not be simply transplanted</li> </ol>
Brown and Pyers	1988			How to upgrade financial reporting in local and state government?	What are the potential benefits?	<ol style="list-style-type: none"> <li>a. The inclusion of performance measures in the financial reporting should enhance the value of these reports for the citizenry</li> <li>b. Behavioral issues will be the barriers, not technical ones</li> </ol>
Carter	1991	What organizational characteristics influence the implementation of performance indicators?		What are the problems of performance assessment and the methods adopted for the design?		<ol style="list-style-type: none"> <li>a. Key organizational characteristics are: ownership, trading status, degree of competition, degree of political accountability, heterogeneity (different products), complexity (different skills) and uncertainty (of objectives, means and ends)</li> <li>b. The private (private monopolies with government control)/ public divide does not explain variation in the problems of performance measurement</li> <li>c. Ideal indicator systems are parsimonious, timely and custom built</li> </ol>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Chalos, P. and Cherian, J.	1995		Are SEA data useful for an educational fiscal policy? Is there an association between educational effectiveness and the source of educational financing (case 1) and the referenda outcomes (case 2)?			a. The results of the DEA analysis of the SEA data have both operational and public policy implications - they may be used in setting budgetary goals and providing feedback with respect to resource utilization relative to other comparable units - the results support the existence of fiscal inequity in education - the lack of a statistically significant relationship between effectiveness/efficiency and referenda outcomes suggests that districts do not recognize the marginal utility of marginal taxation.
Coe, Charles	2004			What is the validity and comparability of scorecards?		a. Assessment of the validity by looking at the face validity and the application and exploitation of weights / assessment of the comparability in current year among units and in multiple year b. Seven recommendations:- invest upfront- be sensitive to trade-offs- comprehensible- control for variables- adjust changes- apply and state weights- join with academicsc. c. Conclusion- there are different purposes of report cards : commercial, influencing policy, academic achievement, competition between scorecards- further research should assess the effects of report cards
Dawson and Street	2000			How is unit cost measured with indices and what is the reliability?	What are the unintended consequences of the publication of cost information?	a. There are considerable variations in ranks across NHS cost indices b. The incentives that the indices create are not always consistent with government objectives c. The measurement of unit cost leads to 'creative accounting'
de Lancer Julnes	2001			Does participation in indicator development increase the perception of usefulness?		The participatory approach does improve perception of usefulness, especially in organizations that feel vulnerable about the effect of performance measurement.

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
de Lancer Julnes and Holzer	2001	What are the factors affecting adoption and implementation of performance measures? What are the factors affecting implementation of performance measures?				<ul style="list-style-type: none"> <li>a. Adoption of performance measures is more heavily influenced by rational/technical factors while implementation is more heavily influenced by political/cultural factors</li> <li>b. Rational/technical factors are having a goal orientation, committing resources, being informed about the techniques and the existence of external requirements</li> <li>c. Political/cultural factors are the existence of internal and external interest groups, and risk taking attitudes in the organization</li> </ul>
England and Parle	1987		What are the performance appraisal practices for non-managerial performance and how does it differ from managerial performance appraisal?			<p>This study repeats the framework of Ammons (1985) on non-managerial performance appraisal and compares the results:</p> <ul style="list-style-type: none"> <li>a. Most city governments have formal appraisal systems in place</li> <li>b. Non-managerial appraisal is more formally documented than managerial evaluation</li> <li>c. Identification of skill deficiencies is more often an objective for non-managerial appraisal</li> <li>d. Rating scales are more often used in non-managerial appraisal (while MBO is less popular),</li> <li>e. There is some variation in techniques when controlling for city size, which is not the case for managerial performance appraisal</li> </ul>
Glaser, Marc	1991			What are the most prominent design and implementation issues of performance measurement in order to make measurement usable?		<ul style="list-style-type: none"> <li>a. Triangulation: ensuring validity of performance measures</li> <li>b. Target setting is not done right and often not ambitious enough</li> <li>c. Performance measurement deals with present and past and is therefore not linked to strategic planning exercises</li> <li>d. Performance measurement is not adequately integrated in the budget process</li> <li>e. Often there is too much inflexibility in the format and design of presentation formats</li> <li>f. Especially the uncertainty about the content of performance measurement invokes resistance</li> </ul>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Goddard and Mannion	2004		What are the main differences in use between the vertical and horizontal approach?	What are the main differences in measurement and analysis between the vertical and horizontal approach?	What are the advantages and disadvantages of the approaches?	<p>a. There is a vertical (top-down) and a horizontal (bottom-up) approach to measurement</p> <p>b. The two approaches have a different measurement, analysis and action design</p> <p>c. Different sectors have elements of both approaches, in different settings</p> <p>d. The delivery of key objectives appears to benefit from a vertical approach; key objective improve measured results</p> <p>e. The dysfunctional consequences arise from a focus on central targets</p> <p>f. Learning benefits from a horizontal approach</p> <p>g. Inspections may reinforce vertical or horizontal approaches</p>
Greener, Ian	2003		What is the current use of performance measurement in the NHS? How to make performance measurement more useful and avoid perverse effects?		What are the effects on the local health trusts?	<p>a. Performance measurement is used as an instrument for control and cost saving by the central government.</p> <p>b. This leads to a redundancy fallacy: - internal markets require redundancy. Purchasers need to be able to switch between providers. - the capacity of NHS managers to absorb change requires redundancy.</p> <p>c. This leads to the fallacy that we can work out, and measure, exactly what we mean by performance.</p> <p>d. The increased use of league tables aggravates the pressure. The author pleads for a decentralization of the evaluation of performance to the professionals.</p>
Grizzle and Pettijohn	2002	Which factors affect the implementation of performance based program budgeting?				<p>a. Critical factors in four categories are identified; communication, resources, dispositions and bureaucratic structure (based on Edwards (1980)).</p> <p>b. These factors are brought together in a systems dynamic model based on case studies in Florida.</p> <p>c. The next step is to test the model and the causal loops empirically.</p>



Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Grizzle, Gloria	1987			How can budget officers play a role in linking performance information to a budget?		budget officers can play a role in linking performance information: a. Workload shedding: computerization; biannual budgets, rotate program review (only new/changed programs; more leeway for agencies to transfer funds) b. Generating efficiency information: invest in technology, budget guidelines to require unit cost, use productivity information for decisions c. Generating outcome information: ask for outcome information selectively, tie evaluation requirements to pilot projects, attach evaluation requirements to appropriations, performance agreements, acquire legislative staff evaluations, enlist graduate students to do evaluations d. Cost effectiveness information: convert outcome evaluations
Grizzle, Gloria	2002		How to design measurement systems that minimize unintended consequences?		What are the unintended consequences?	1, Unintended consequences of outcome measurement: teaching for the test, systematic downgrading of offences, distorting accounting records of revenues, creaming 2. Unintended consequences of customer satisfaction; guiding the answers 3. Unintended consequences of Measuring quantity of work performed; definition creep, lowering quality 4, Unintended consequences of efficiency measures: decentralization of administrative functions 5. Remedies a. General standards of integrity b. More rules and supervision 6. An alternative is to rely on trust and professionalism instead

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Halachmi, Arie	2002				Does performance measurement facilitate better performance? Does performance measurement increase accountability? Can performance measurement increase both?	<ol style="list-style-type: none"> <li>1. Performance measurement siphons resources from "production" to overhead</li> <li>2. These costs are justified in the name of two important values: the need to assure accountability and the need to improve performance.</li> <li>3. The accountability motive has to do with continuing mistrust of government operatives</li> <li>4. The intellectual roots of the improved performance motive are Taylor and PA theories</li> <li>5. Accountability improvement may be inconsistent with the effort to enhance performance, since the first objective requires synthesis, intuition and creativity and the second requires rational, systemic analysis</li> </ol>
Hatry, Harry P.	2002		What are the fashions and fallacies in performance measurement and management?	What are the fashions and fallacies in performance measurement and management?		<ol style="list-style-type: none"> <li>a. Performance measurement is not the same as performance management</li> <li>b. Strategic planning needs to use performance measurement</li> <li>c. Performance measurement looks back, budgeting and policy planning looks forward</li> <li>d. Outcome estimates are seldom a sound basis for budget decisions</li> <li>e. Outcomes can seldom be readily linked to input requirements</li> <li>f. Customer surveys need not always survey customer satisfaction</li> <li>g. Surveys do not always need high response and precision</li> <li>h. Aggregate information needs to be broken out</li> <li>i. The primary purpose of performance information is to improve services, not only accountability</li> <li>j. Public officials cannot be held fully accountable for results</li> </ol>
Hedley	1998	Can private sector methods be used for public sector performance measurement?				<ol style="list-style-type: none"> <li>a. Employing private sector standards is replete with problems, the most significant of which is the seemingly incompatibility of efficiency and effectiveness controls.</li> <li>b. The standards and expectations for public sector output are different and more complex.</li> </ol>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Heinrich	1999		Do job-training agency's program administrators use the performance standards system?		Do the performance standards motivate contractor behavior that it enhances agency performance consistent with program goals?	a. Information from the performance based contracting system is being used to make resource allocation decisions.b. performance measures are not strongly correlated with program goals.c. there is a predominance of cost-per-placement considerations which has a negative impact on service quality.
Henry and Dickey	1993			Can an R&D approach be useful for a performance monitoring system?		An R&D framework can be useful in guiding the choices that an organization has to make in setting up a performance monitoring system.
Ho and Coates	2004				What are the results of the citizen initiated performance assessment project in Iowa?	a. Impact on departments: departments used citizen involvement for new tools and practices b. Impact on policies and procedures; cities implement procedures for citizen involvement c. Impact on performance measurement; different indicators are proposed d. Impact on city staff perspective; new perspectives on the citizen e. Challenge; educate citizens on the role of PI's (PMsmt is not policy making) f. Challenge; how to sustain interest? g. Challenge; how to involve young and minority population? h. Challenge; how to overcome staff resistance?
Ho, Alfred	2003	Is mandatory performance reporting (attached to GAAP) feasible in small cities?				a. There is a generally positive evaluation of the concept b. Practical barriers such as data availability are less of a concern c. There is a lack of communication between administrators and elected officials

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Hyndman and Anderson	1995			To what extent do agencies (U.K.) report performance information?		<p>a. 42% of the agencies reported no performance information on efficiency</p> <p>b. 14% reported no performance information on effectiveness</p> <p>c. There is a lack of early formal guidance on how to measure and to report performance information.</p> <p>d. The change in focus shifts from the component parts of performance (i.e. input, output) towards higher measures of performance (i.e. efficiency and effectiveness)</p>
Ingraham, P	1993	<p>a. to what extent is PFP successful in the private sector (as perceived by public policy makers)?</p> <p>b. to what extent is PFP in the private sector transferable to the PS?</p> <p>c. to what extent do the conditions for effective PFP in the PS exist?</p>	How can it be implemented in the PS, given the typical conditions?			<p>a. Evidence of success in the private sector is much more limited than generally believed.</p> <p>b. The context of the private sector is fundamentally different: civil service laws and procedures limit discretion, line and personnel management are decoupled, and limited financial resources are devoted to award performance.</p> <p>c. The transfer of PFP would thus require a fundamental revision of the management structure.</p> <p>d. In order to implement PFP in the PS, we need:</p> <ul style="list-style-type: none"> <li>&gt; to look at what organizations need</li> <li>&gt; to find examples in the public sector</li> <li>&gt; to look at threshold conditions for success</li> <li>&gt; to fit PFP into the civil service system</li> <li>&gt; to provide politicians and top managers with information on PFP</li> <li>&gt; to look at other pay benefit schemes</li> <li>&gt; to research the phenomenon</li> </ul>
Johnsen	1999			What is the impact of the implementation mode (coupled to objectives or not) for introducing performance in Norwegian local government?		<p>a. The successful cases used a decoupled implementation mode, i.e. by not linking performance indicators to objectives; decoupling may lead to an increased instrumental use of performance information.</p> <p>b. This may be explained by ambiguity; implementation requires power, which requires coalitions and 'overselling' of the project. PI's that are tightly coupled to oversold goals may be unreliable and vague.</p> <p>c. Another explanation may be an increasing environmental resistance and conflict over the goals, if coupled with indicators</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Kravchuk and Schack	1996			What are the challenges in designing more formal performance measurement systems under GPRA?		Ten design principles are proposed; a. Formulate a clear mission, strategy and objectives; b. Develop an explicit measurement strategy; c. Involve key users; d. Rationalize the programmatic structure as a prelude to measurement; e. Develop different sets of measures for different users; f. Consider the customers of the program; g. Provide each user with sufficient detail; h. Review and revise periodically the performance measurement system; i. Take account of upstream, downstream and lateral complexities; j. Avoid excessive aggregation of data. Finally, an additional eleventh point is to recognize and use measures as "indicators" only.
Lawton, A., McKeivitt, D. and Millar, M.	2000	What is the source of impetus behind the development of performance measurement systems?		What is the role played by major stakeholders in the development of PMS?  What is the extent of user involvement in developing measures of effectiveness and quality of service delivery?		a. The authors use institutional theory to study performance measurement systems b. The study (a survey of middle management participants to an MBA program) showed that - the implementation of performance measurement has a top-down character, - that there is no linkage between the impetus and the operational change - that there is no attention to the views of the client in the process of performance measurement and management. c. additionally, the authors find that middle managers feel distanced from performance measures which are devised without their active participation.

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Lee and Burns	2000	What are the explanations for changes in the use of performance information in budgeting?	What was the status of performance measurement in state budgeting as of 1995? What changes occurred between 1995 and 1990 in the use of performance measurement in state budgeting?			<p>a. The integration of performance information in the budget follows the ebb and flow metaphor rather than a one way street</p> <p>b. The cause is not determined in this study (% of urban population, non agricultural employment rate, per capita income, tax capacity, tax effort and the state's unemployment rate did not correlate significantly)</p> <p>c. Larger states were less likely to backslide than smaller states</p>
Lindkvist	1996	How may the failure of performance based budgeting be explained?			What are the results of the introduction of a performance based compensation system in a Swedish university hospital?	<p>a. Performance based compensation did not produce the desired results. It is problematic to introduce a market oriented budget system in health care</p> <p>b. Causes are the limited measurability of the services, the existence of local monopolies and the lack of knowledge on prices and costs</p> <p>c. Reforms in healthcare and reliance on economist reasoning should align with the image of the firm; in this case the image consists of a long-term knowledge developing organization that is driven more by professional than economic incentives.</p>
Mausloff	2004		How can performance information be used for learning and feedback?			<p>a. Learning has four phases; identifying the performance gap, integrating the views within the organization, searching (focused or scanning) for solutions and implementation (anew theory of action)</p> <p>b. These phases are not subsequent, phases are skipped and have different sequences.</p> <p>c. Stress is a positive but unnecessary condition for learning</p> <p>d. Every phase yields new information, and thus new interpretations. This is a another concept than traditional organizational learning literature.</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Mayston	1985		What roles do non-profit performance indicators play in the public sector?			<p>role 1; Clarifying the organization's objectives</p> <p>role 2; Evaluating the final outcomes resulting from an organization's activities</p> <p>role 3; Providing an input to managerial incentive schemes</p> <p>role 4; Enabling consumers to make informed choices</p> <p>role 5; Introducing performance standards in licensing or contracting of privatized services, and monitor fulfillment</p> <p>role 6; Indicating the effectiveness with which different service activities in a given policy area contribute to relevant performance dimensions</p> <p>role 7; Being a trigger for further investigation and remedial action</p> <p>role 8; Assisting in determining the most cost effective set of service levels</p> <p>role 9; Indicating areas of potential cost saving</p>
McKevitt and Lawton	1996		Is PM fulfilling the institutional function (meeting expectations from external stakeholders), the managerial function (professionals and managers), and the technical function (customers and clients)?			<p>a. PM has been used as a top-down instrument of senior management control.</p> <p>b. PM was implemented to satisfy demands of the institutional environment</p> <p>c. Professional concerns feature prominently in NHS PM</p> <p>d. There is a lack of attention for middle and junior management.</p> <p>e. There is a lack of attention for the user of the services</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Melkers and Willoughby	1998		How many states have adopted some version of Performance Based Budgeting?			<p>a. 47/50 states have PPB requirements:</p> <ul style="list-style-type: none"> <li>- 16/47 states only have non-legislative requirements</li> <li>- 31/47 also have legislative requirements.</li> </ul> <p>b. 16/31 link performance measures with strategic planning in law.</p> <p>c. Only 7 states establish positive incentives for achieving goals in their legislation, whereas two states have negative incentives for not achieving the goals.</p> <p>d. Only few states created offices to assist in the performance measurement process.</p>
Melkers and Willoughby	2001	What was the impetus for the performance budgeting reform?	To what extent are the reforms implemented?		What were the effects of PBB?	<p>a. Budgeters attributed an improved decision making to performance budgeting initiatives.</p> <p>b. Implementation however is still challenging: resources need to be available and organizational cultures need to change.</p> <p>c. Few states indicate any link between performance information and actual appropriations. This is not alarming, since the reform is time consuming.</p>
Modell	2004				How do changes in PM gradually come to be implicated in the social fabric of organizations or fail to achieve the status of widely accepted myths?	<p>a. A dominant myth is challenged by a crisis, a new ghost myth is formed and institutionalized, the new myth takes over, although some remnants of the old myths may remain.</p> <p>b. The '80s PM myth was about financial control, the '90s control myth is about goal directed and multidimensional, and shared models.</p> <p>c. PM myth pivoting around the supremacy of goal-directed, multidimensional PM models may gradually replace the myth that public service provision may be improved by heavy reliance on financial control.</p>



Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Mol, Nico P.	1996			Which are the problems with the construction and use of relevant indicators for performance evaluation?		Indicators for contract management program have several deficiencies which need to be addressed: <ul style="list-style-type: none"> <li>- a lack of relevance. The majority of the indicators refer to secondary processes.</li> <li>- a lack of coherence. Indicators are not related to each other.</li> <li>- a lack of consistency (especially in relating input to output and effect)</li> <li>- a lack of completeness (especially for what budget information is concerned)</li> <li>- the performance controls are not substituting the input controls.</li> </ul>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Moynihan, Donald P. and Ingraham, Patricia W.	2003	Why do states achieve high grades on management capacity?		Which management practices are designed to enable MFR reforms to work as intended: the integrative facilitators?		<p>a. Significant governance factors include the maturity of the state (+), population and diversity (+), the number (+) and diversity (-) of the interest groups, social capital, the entrepreneurial climate (+), and the professionalism of the legislature (+)</p> <p>b. Established economic controls such as income, income equality and education have little support</p> <p>c. Management For Results works as intended:</p> <ul style="list-style-type: none"> <li>- when there is strategic planning with clear purposes, communication, and coordination</li> <li>- when indicators and evaluative data are developed to monitor progress and are valid and accurate</li> <li>- when the data is being used in decision venues</li> <li>- when government clearly communicates results to stakeholders</li> </ul> <p>d. The integrative facilitators are:</p> <ul style="list-style-type: none"> <li>- the comprehensiveness of the MFR system throughout government at both the statewide and the agency levels, which requires a common framework and a common language.</li> <li>- the vertical integration of goals among the center, the agencies and the programs</li> <li>- the balancing of top-down and bottom-up approaches</li> <li>- the provision of clear guidance for agency efforts: standardization of types of information. (Central agencies play an important facilitating role, expertise is largely stored in-house although often consultants are involved)</li> <li>- leadership and commitment in both the legislative and the administrative branches</li> </ul>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Nyhan and Martin	1999			What are the possibilities and restrictions of the application of DEA for comparative performance measurement?		a. Performance measurement will become increasingly important due to GPRA, NPR, community benchmarking and GASB's SEAb. Improved analytical tools such as Data Envelopment Analysis in addition to ratio analysis and regression analysis may be useful
O'Toole, D. and Stipak, Brian	1988	What types of local government organizations are the most productivity oriented? How does the budgeting environment toward productivity improvement efforts evolve?	To what extent are productivity measures used in budget practices?			a. Different types of performance measures have widespread use and influence, there are more sophisticated financial forecasting methods available and output oriented budget formats are increasingly used b. A higher population and budget leads to higher values on productivity orientation, the existence of budget office leads to more productivity efforts
Poister and Streib	1989		What are the uses of management tools in local government?		What is the effectiveness of the tools?	a. There is a widespread, enduring use of management tools, amongst which performance monitoring b. Effectiveness is evaluated positively
Poister and Streib	1999	What is the motivation of the most accomplished practitioners?	What is the extent to which performance management has become integrated into contemporary local government management?			a. 40% make any kind of meaning full use of performance measures b. Programs are seldom comprised in comprehensive systems that already exist c. Cities do more often use workload or effectiveness measures than unit cost or efficiency measures d. Motivation is intern (better decision making) and not so much extern (legal requirements) e. There are often ownership problems with management (45%) and rank-in-file (60%), but less with the council (30%)

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Reck	2001	What is the influence of resource scarcity on the association between financial information, non financial information and resources allocation decisions?	To what extent is non financial and financial performance information used? Is there incremental value in adding non financial information to financial performance information?			<p>a. Non-financial information does not play a role in resource allocation decisions, financial information does</p> <p>b. Non financial information however is significantly associated with the respondent's evaluation of unit and individual performance</p> <p>c. When resources are scarce, the association between financial information and budget allocation of resources declines</p>
Rivenbark and pizzarella	2002			<p>Why audit performance data?</p> <p>How to audit performance data?</p>		<p>a. The following areas affect performance data: complexity of processes, organizational changes, interpretation of measures, reporting capabilities, functional boundaries</p> <p>b. The proposed audit approach consists of ten steps: audit schedule, entrance conference, scope of audit, methodology, findings, recommendations, implementation guidelines, exit conference, audit report and follow-up interview</p>
Roy and Séguin	2000	Why do organizations in the public sector adopt efficiency oriented approaches?				Organizations adopt efficiency oriented approaches because of institutional pressure. This research confirms Powell and Dimaggio's institutional isomorphism theory.

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Rubenstein, Schwartz and Stiefel	2003			How can organizational performance be accurately assessed when the contextual factors are varied and the effects of these factors on the output are not completely understood?		a. The technique of the adjusted performance measures can incorporate environmental factors into the analysis. In essence, this is a regression analysis.
Rutherford	2000			How does the construction and presentation of PI's affect understandability, comparability and user's perceptions of their importance?		<p>a. Construction and presentation of performance indicators are important for understandability, comparability and the perception of importance</p> <p>b. The commentary on the performance indicators is highly variable and often very poor</p> <p>c. The terminology was certainly not as fully standardized as the Treasury and the Civil Service Committee expected</p> <p>d. The potential for comparisons in time with indicators is missed due to flexibility in shaping the indicators that is allowed to the agencies</p> <p>e. Indices may increase understandability, but they are seldom used and have serious pitfalls</p> <p>f. Central government should come to a meaningful standardization (and thus a maximum comparability) without imposing a 'one size fits all' framework</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Sanderson, I.	2001	What are the conditions for a successful implementation of the Best Value framework in the UK local governments?	What is the rationale for performance evaluation: control or evaluation?			<p>a. The rationale of performance management under the Conservatives was control and 'upwards accountability' while the rationale of Labour's Best Value framework is continuous improvement</p> <p>b. Requirements are</p> <ul style="list-style-type: none"> <li>- the development of the capacity to achieve change</li> <li>- the consistency of the nature of change with the key goals of the local authorities</li> <li>- the involvement of all major stakeholders</li> </ul> <p>c. These requirements imply a change in the organizational culture and a supportive infrastructure</p>
Schneider, Martin	2004		How to manage performance of complex duties, in the absence of monetary incentives and effective sanctions?			<p>a. In this professional context, performance management can rely on maintaining and directing norms and values. Performance management by culture</p> <p>b. This cultural approach combines elements of professional ethics and bureaucratic control:</p> <ul style="list-style-type: none"> <li>- through peer review</li> <li>- through internal recruitment</li> <li>- through soft benchmarking</li> </ul> <p>c. The similarities between USA and Germany are notable, notwithstanding institutional differences</p> <p>d. High-powered incentives are probably not effective.</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Serban and Burke	1998			what are the main phases and elements of (the supply of information for) performance funding ?		<ul style="list-style-type: none"> <li>a. Public accountability and institutional improvement should be considered as complementary purposes</li> <li>b. State prescription of indicators seldom works, but a general mandate can be important for action</li> <li>c. State-wide coordinating agencies are important</li> <li>d. Performance indicators should be a means for improvement, and not an end</li> <li>e. The amount of performance funding should be large enough to spur institutional improvement, but not so large as to destabilize institutional budgets</li> <li>f. The fairest approach is institutions competing to improve past performances</li> <li>g. Performance measurement needs to be made visible and meaningful to deans and faculty</li> </ul>
Smith	1988				Did the conservative government's (installed in 1979) policy to increase competition actually lead to an increased competition amongst local councils?	<ul style="list-style-type: none"> <li>a. The various initiatives to increase competition (among other things the reporting of performance indicators, the use of block grants, rate capping, the targets regime and the abolition of the urban strategic authorities) did not lead to an increased competition</li> <li>b. In order to come to this conclusion, the variation of indicators that councils report is used as a proxy for competition, where reduction in variability is considered to reflect increased competition, assuming that there is a 'model' pattern of service delivery to which groups of authorities want to converge.</li> </ul>
Smith	1990		What is the purpose of public sector information systems? What are the developments in England and Wales to 1989?	What are the techniques for measuring performance?		<ul style="list-style-type: none"> <li>a. The main purpose is accountability, which is more problematic in the public than in the private sector.</li> <li>b. Main applications are in local government, health.</li> <li>c. Techniques of analysis are ratio analysis, cluster analysis, regression analysis, and DEA</li> </ul>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Smith	1993		What is the rationale for the dissemination of outcome performance data to secure enhanced strategic control?		What are the potential distortions?	<p>a. The rationale for outcome related performance indicators is both intern control and external accountability</p> <p>b. Outcome reporting schemes are embedded in an organization, and thus are not neutral</p> <p>c. An over-reliance on performance indicators for control has dysfunctional consequences; tunnel vision, sub-optimization, myopia, convergence, gaming, ossification, misrepresentation</p>
Ter Bogt, H.	2004		Do aldermen use performance information?			<p>a. Aldermen prefer rich, verbal information to sources of written information</p> <p>b. Aldermen are hardly involved in the production of much of the information</p> <p>c. No significant differences were found between the different portfolio groups</p> <p>d. The performance measurement is hardly used for organizational controle. there is no evidence for isomorphism</p>
Thompson	2000				What are the consequences of the use of performance measures on management practices for mid and lower level management? Does total management capacity decrease?	<p>a. The Social Security Administration has a strong use of the "big four" indicators (output indicators on processing of claims). These indicators are used for pay decisions, promotions, and ranking district offices with a lot of gaming as a consequence.</p> <p>b. Measures became a substitute for a more in depth decision making process.</p> <p>c. Measures have been used by top management and political levels to increase control over middle and low level management</p> <p>d. By using the measures for control purposes, the discretion needed to implement the management function at mid- and lower level has eroded. The total management capacity is reduced.</p> <p>e. Because of the fact that management functions are intertwined, the affection of one management function results in a more than proportional reduction in management capacity at mid- an lower levels.</p> <p>f. There is a dual potentiality in performance measurement, control and improvement, which may not easily be combined</p>



Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Van Thiel and Leeuw	2002				<p>What are the intended and unintended effects of performance measurement in the public sector?</p> <p>How can these effects be detected?</p>	<p>a. NPM led to an increased attention for performance assessment. However, there are unintended consequences such as increased monitoring costs, dysfunctional effects, symbolic behavior, and unclarity about what is being measured. One unintended consequence is studied more in depth; the performance paradox.</p> <p>b. The performance paradox states that there is a weak correlation between performance indicators and performance. This is the result of four processes.</p> <ul style="list-style-type: none"> <li>- positive learning</li> <li>- perverse learning</li> <li>- selection</li> <li>- suppression</li> </ul> <p>c. These processes occur when there are minimal accountability requirements, when the policy objectives are elusive, when policy goals are hard to measure, and when there is a strong emphasis on monitoring in the organization.</p> <p>d. A deliberate performance paradox may be the result of a discrepancy between the politician's and agency head's goals, the lack of a potential bankruptcy, and the disjunction of costs and revenues</p> <p>e. The detection of a performance paradox ex ante is difficult. However, a low number of indicators, indicators that are developed by the organization only, gaps in performance reports and the occurrence of guidelines to deal with audit, may point to a performance paradox.</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Wang	2002		What is the influence of implementation strategies on PM impact?	Does PM have any impact in U.S. city governments?		<p>a. The impact of performance measurement is substantial for internal management, but very limited in the communication and accountability with stakeholders (legislators, the public, chief executives) and resource allocation decisions.</p> <p>b. In order to involve citizens, the author proposes to develop quality of life indicators</p> <p>c. The impact of the tool 'performance measurement' may be sustained if it is linked with a purpose.</p> <p>d. Strategic planning increases the impact and may thus be the most apt purpose for performance measurement in the public sector.</p>
Wang and Berman 2000	2000	What is the impact of organizational relationships, structures and goals on the deployment of performance measurement?		What is the extent and the nature of performance measurement in counties in the USA?		<p>a. The most important independent variables were external support from elected officials and citizens, mission orientation and central management involvement</p> <p>b. Professional competency and resource availability are also associated with the presence of performance measurement, however, less importantly</p> <p>c. Decentralization efforts are not associated</p>
Wholey and Hatry	1992		Is performance monitoring worthwhile?	Is performance monitoring in public programs feasible?		<p>a. Performance monitoring is feasible, several organizations already do itb. performance monitoring is necessary for implementing modern management techniquesc. barriers to measurement are six fold:</p> <ul style="list-style-type: none"> <li>&gt; outcome versus impact: causality?</li> <li>&gt; validity and reliability?</li> <li>&gt; what are acceptable performance levels?</li> <li>&gt; cost</li> <li>&gt; reporting fears</li> <li>&gt; lack of utility for the program manager; timeliness and over aggregation.</li> </ul>
Wiggins and Tymms	2002				What are the unintended or dysfunctional effects of high-stakes single proxy indicators?	English primary schools perceive their KPI systems (with league tables) as being significantly more dysfunctional than those of their Scottish counterparts

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Williams, Daniel W.	2003	What was the context of the initiative?	How did the NY Bureau of Municipal Research (BMR) use performance data?	How did the BMR gather data?		<p>a. - The <b>context</b> was the shifting power to a singular executive model and the commission model with executive power in the hands of the legislature. The shift was caused by an increase of the importance of the city administration combined with corruption, distrust of the electorate and a wish to consolidate the executive branch by eliminating numerous administrative positions. Performance measurement was used to reconcile these conflicting views.</p> <p>- Four factors explained the success of the BMR:</p> <ul style="list-style-type: none"> <li>(i) connected to the academic establishment</li> <li>(ii) intentional exportation of the work to other communities</li> <li>(iii) direct contact with government agencies</li> <li>(iv) promotion through the federal government</li> </ul> <p>b. The information was <b>used</b> for reporting purposes, as for budget allocations and productivity improvement</p> <p>c. The <b>data gathering</b> relied on the accounting system as record keeping. Input, output and outcome indicators were pursued, as social indicators and needs assessment.</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Williams, Daniel W.	2004		What is the evolution of performance measurement?			<p>a. Performance measurement as developed by the NY bureau for Municipal Research has three ancestors.</p> <ul style="list-style-type: none"> <li>- the social survey of the settlement movement, inspired by Charles booth's poverty mapping</li> <li>- municipal statistics</li> <li>- cost accounting</li> </ul> <p>b. Four developments took place after the initial initiatives of the NYBMR</p> <ul style="list-style-type: none"> <li>- research into government management (implementing a functional budget and accounting system, standardization of work processes by personnel</li> <li>- development of surveys (of the entire community, of special topics and of government issues)</li> <li>- development of scorecards, and later on indexes</li> <li>- increasing specialization</li> </ul> <p>c. Key changes are</p> <ul style="list-style-type: none"> <li>- an increased sophistication</li> <li>- a higher quantification</li> <li>- a focus on results</li> <li>- a decline in activism: value neutral science ( influenced by the shift towards the executive)</li> <li>- a disinterest in communication with the public</li> <li>- a more narrow focus on government instead of governance</li> </ul> <p>d. Conclusion: performance measurement does not refer to a particular technique. Instead, it refers to he application of relevant techniques to the problem of observing government at work.</p>
Wilson	2004			Are the school rankings in the UK robust? Is the value added PI a good indicator?		<p>a. There is no relationship between the target PI and the value added PI</p> <p>b. Value added PI's take the intake into account</p> <p>c. School rankings are sensitive to the PI employed</p>

Author	Date	Causes conditions	Demand	Supply	Results, Effects	Conclusions
Woodbury and Dollery	2003			Is the current performance measurement regime in Australia sufficiently developed to enable policymakers to evaluate reforms?		a. Data production problems and problems with techniques are prevalent in Australian local government. b. A broad application of DEA would yield the best results for assessing efficiency of local governments
Worthington and Dollery	2000			How can efficiency of local governments be evaluated using a nonparametric approach (DEA)		a. Technical and scale efficiency differs among the councils b. Technical inefficiency contributes more than scale inefficiency to total differences c. Even when controlling for external factors, differences remained. These techniques may be integrated in intergovernmental grant systems

### 13.1.3. Classification of the studies: focus, theoretical orientation, theoretical range and method

Author	Date	Focus 1	Focus 2	Focus 3	Orientation	Range	Method I	Method II
Alcock, Peter	2004	administration	politicians		positivist	Middle Range	obtrusive small N	
Ammons and Rodriguez	1986	administration			positivist	Small Range	obtrusive large N	
Ammons, Coe, and Lombardo	2001	administration			positivist	Small Range	obtrusive small N	
Ammons, D.N	1985	administration	politicians	citizens and press	positivist	Small Range	obtrusive small N	
Ammons, D.N	2002	administration			positivist	Middle Range	literature study	
Andrews	2004	administration	politicians		positivist	Small Range	unobtrusive large N	
Behn	2003	administration	politicians	citizens and press	positivist	Middle Range	literature study	
Behn	2002	administration	politicians	citizens and press	positivist	Middle Range	literature study	
Behn and Kant	1999	administration			positivist	Middle Range	literature study	
Berman and Wang	2000	administration			positivist	Middle Range	obtrusive large N	
Berman, Evan	2002	administration			positivist	Small Range	literature study	
Berry, Brower and Flowers	2000	administration	politicians	citizens and press	interpretative	Middle Range	unobtrusive small N	unobtrusive large N obtrusive small N
Bouckaert	1993	administration			positivist	Middle Range	literature study	
Bouckaert and Peters	2002	administration			positivist	Middle Range	literature study	
Bouckaert, Geert	1990	administration			positivist	Historical	literature study	
Boyne and Law	1991	administration	politicians	citizens and press	positivist	Small Range	unobtrusive large N	
Broom	1995	administration	politicians		positivist	Small Range	obtrusive small N	
Brown and Pyers	1988	administration	politicians	citizens and press	positivist	Middle Range	obtrusive small N	
Carter	1991	administration			positivist	Middle Range	unobtrusive small N	
Chalos, P. and Cherian, J.	1995	administration			positivist	Small Range	unobtrusive small N	
Coe, Charles	2004	administration	politicians		positivist	Small Range	unobtrusive small N	
Dawson and Street	2000	administration	politicians		positivist	Small Range	unobtrusive small N	
de Lancer Julnes	2001	administration	politicians		positivist	Small Range	obtrusive small N	
de Lancer Julnes and Holzer	2001	administration	politicians		interpretative	Middle Range	obtrusive large N	
England and Parle	1987	administration			positivist	Small Range	obtrusive large N	
Glaser, Marc	1991	administration			positivist	Small Range	obtrusive small N	
Goddard and Mannion	2004	administration			positivist	Middle Range	unobtrusive small N	
Greener, Ian	2003	administration	politicians		positivist	Middle Range	obtrusive small N	

Author	Date	Focus 1	Focus 2	Focus 3	Orientation	Range	Method I	Method II
Grizzle and Pettijohn	2002	administration	politicians	citizens and press	interpretative	Middle Range	obtrusive small N	
Grizzle, Gloria	1987	administration			positivist	Small Range	literature study	
Grizzle, Gloria	2002	administration			positivist	Small Range	literature study	
Halachmi, Arie	2002	administration	politicians		positivist	Middle Range	literature study	
Hatry, Harry P.	2002	administration			positivist	Small Range	literature study	
Hedley	1998	administration			positivist	Small Range	obtrusive small N	
Heinrich	1999	administration			positivist	Middle Range	obtrusive small N	
Henry and Dickey	1993	administration			positivist	Small Range	obtrusive small N	
Ho and Coates	2004	administration	politicians	citizens and press	positivist	Small Range	obtrusive small N	
Ho, Alfred	2003	administration	politicians		positivist	Small Range	obtrusive large N	
Hyndman and Anderson	1995	administration			positivist	Small Range	unobtrusive large N	
Ingraham, P	1993	administration	politicians		positivist	Small Range	obtrusive large N	
Johnsen	1999	administration	politicians		positivist	Middle Range	obtrusive small N	
Kravchuk and Schack	1996	administration			positivist	Small Range	literature study	
Lawton, A., McKeivitt, D. and Millar, M.	2000	administration			positivist	Middle Range	obtrusive small N	unobtrusive small N
Lee and Burns	2000	administration			positivist	Middle Range	obtrusive large N	
Lindkvist	1996	administration			positivist	Middle Range	obtrusive small N	
Mausolff	2004	administration			interpretative	Middle Range	obtrusive small N	
Mayston	1985	administration	politicians	citizens and press	positivist	Small Range	unobtrusive small N	
McKeivitt and Lawton	1996	administration	politicians	citizens and press	positivist	Middle Range	obtrusive small N	unobtrusive small N
Melkers and Willoughby	1998	administration	politicians		positivist	Small Range	obtrusive large N	unobtrusive large N
Melkers and Willoughby	2001	administration			positivist	Small Range	obtrusive large N	
Modell	2004	administration			interpretative	Broad range	literature study	
Mol, Nico P.	1996	administration			positivist	Small Range	obtrusive small N	
Moynihan and Ingraham	2003	administration	politicians	citizens and press	positivist	Middle Range	obtrusive small N	
Nyhan and Martin	1999	administration			positivist	Small Range	literature study	
O'Toole, D. and Stipak, Brian	1988	administration			positivist	Middle Range	obtrusive large N	
Poister and Streib	1989	administration			positivist	Small Range	obtrusive large N	
Poister and Streib	1999	administration	politicians		positivist	Middle Range	obtrusive large N	
Reck	2001	administration			positivist	Middle Range	obtrusive large N	

Author	Date	Focus 1	Focus 2	Focus 3	Orientation	Range	Method I	Method II
Rivenbark and pizzarella	2002	administration			positivist	Small Range	obtrusive small N	
Roy and Séguin	2000	administration			interpretative	Middle Range	obtrusive small N	
Rubenstein, Schwartz and Stiefel	2003	administration			positivist	Small Range	unobtrusive small N	
Rutherford	2000	administration			positivist	Small Range	unobtrusive large N	
Sanderson, I.	2001	administration	politicians		positivist	Small Range	unobtrusive small N	
Schneider, Martin	2004	administration			interpretative	Middle Range	obtrusive small N	
Serban and Burke	1998	administration			positivist	Middle Range	unobtrusive small N	
Smith	1988	administration			positivist	Middle Range	unobtrusive large N	
Smith	1990	administration			positivist	Middle Range	literature study	
Smith	1993	administration			positivist	Middle Range	obtrusive small N	
Ter Bogt, H.	2004	administration	politicians		positivist	Middle Range	unobtrusive large N	
Thompson	2000	administration			positivist	Middle Range	obtrusive small N	
Van Thiel and Leeuw	2002	administration	politicians		interpretative	Middle Range	literature study	
Wang	2002	administration	politicians	citizens and press	positivist	Middle Range	unobtrusive small N	obtrusive large N
Wang and Berman	2000	administration	politicians	citizens and press	positivist	Middle Range	obtrusive large N	
Wholey and Hatry	1992	administration			positivist	Small Range	literature study	
Wiggins and Tymms	2002	administration			positivist	Middle Range	obtrusive large N	
Williams, Daniel W.	2003	administration	politicians		positivist	Historical	literature study	
Williams, Daniel W.	2004	administration	politicians		positivist	Historical	literature study	
Wilson	2004	administration			positivist	Small Range	unobtrusive small N	
Woodbury and Dollery	2003	administration			positivist	Small Range	unobtrusive small N	
Worthington and Dollery	2000	administration			positivist	Small Range	unobtrusive large N	



## 13.2. Descriptive statistics for chapter 6

### A. Adoption (standardized frequencies)

1. What is the coverage rate of the activities?

**STANDCOV**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	25	33,3	33,3	33,3
,33	17	22,7	22,7	56,0
,67	26	34,7	34,7	90,7
1,00	7	9,3	9,3	100,0
Total	75	100,0	100,0	

---

2. Do indicators figure in the annual report of the section?

**annual report**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	43	57,3	57,3	57,3
1,00	32	42,7	42,7	100,0
Total	75	100,0	100,0	

---

3. Does your section apply quality models?

**STAND\_QM**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	48	64,0	64,0	64,0
1,00	27	36,0	36,0	100,0
Total	75	100,0	100,0	

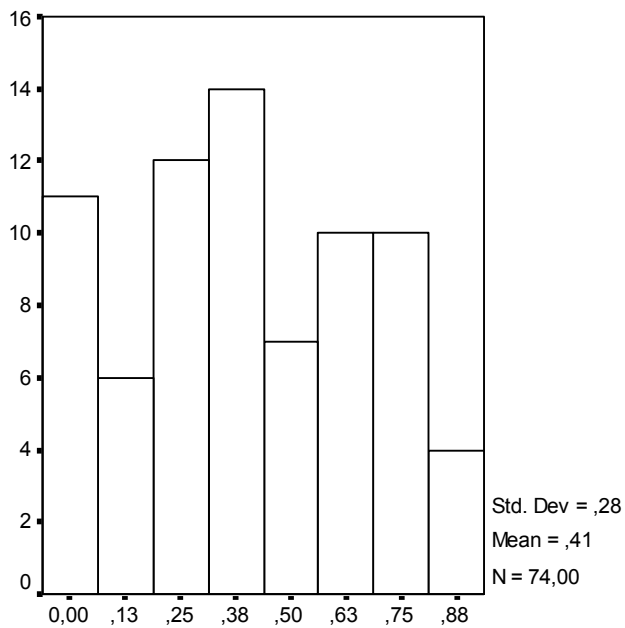
---

4. Is performance measurement an obligation of higher hierarchy?

STAN\_OBL

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	33	44,0	44,6	44,6
	,50	19	25,3	25,7	70,3
	1,00	22	29,3	29,7	100,0
	Total	74	98,7	100,0	
Missing	System	1	1,3		
Total		75	100,0		

5. Adoption Composite Measure



STANDADP

## B. Implementation

1. What is the attendance by top and middle management?

**STD\_PART**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	25	33,3	33,3	33,3
,33	12	16,0	16,0	49,3
,67	25	33,3	33,3	82,7
1,00	13	17,3	17,3	100,0
Total	75	100,0	100,0	

---

2. How often does your section discuss the results?

**STD\_FQD**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	30	40,0	40,0	40,0
,17	10	13,3	13,3	53,3
,33	1	1,3	1,3	54,7
,50	13	17,3	17,3	72,0
,67	15	20,0	20,0	92,0
,83	6	8,0	8,0	100,0
Total	75	100,0	100,0	

---

3. To what extent is the performance information is used to change processes

**STD\_CHGE**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	26	34,7	34,7	34,7
,20	3	4,0	4,0	38,7
,40	6	8,0	8,0	46,7
,60	14	18,7	18,7	65,3
,80	16	21,3	21,3	86,7
1,00	10	13,3	13,3	100,0
Total	75	100,0	100,0	

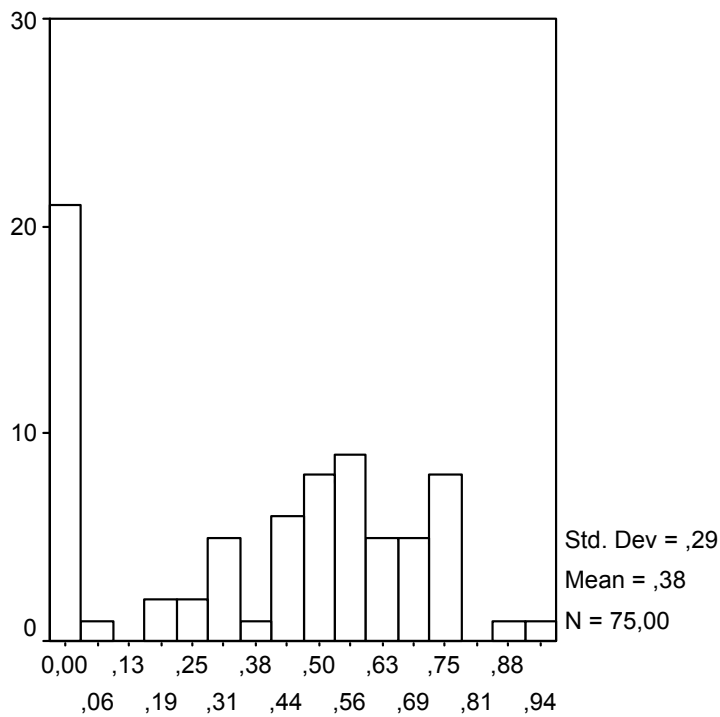
---

4. To what extent is the performance information is used to make allocation decisions?

**STD\_ALLO**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	32	42,7	42,7	42,7
	,20	11	14,7	14,7	57,3
	,40	6	8,0	8,0	65,3
	,60	13	17,3	17,3	82,7
	,80	8	10,7	10,7	93,3
	1,00	5	6,7	6,7	100,0
	Total	75	100,0	100,0	

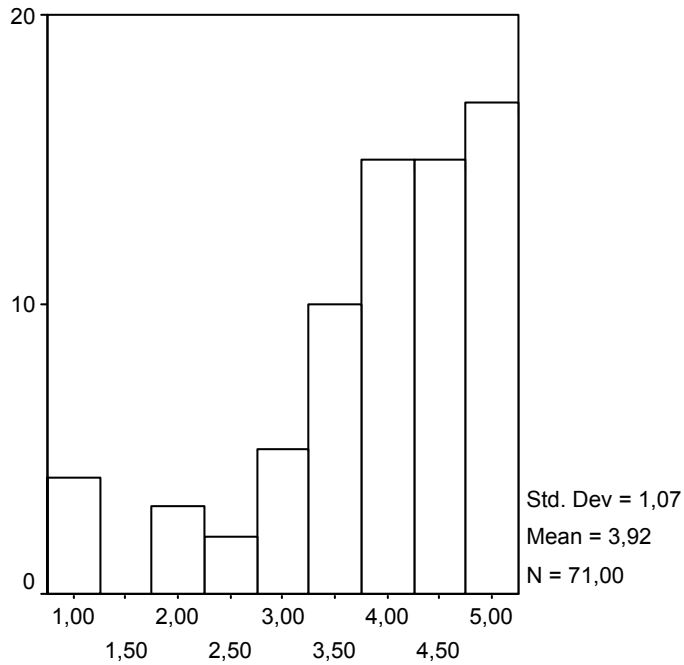
5. Implementation Composite Measure



**STANDIMP**

### C. Descriptive statistics of the explanatory variables

1. Measurability (mean of three scores on a likert-scale for the three main activities of the section)



mean measurability

2. Lack of political involvement is a hindrance (1 not a hindrance to 5 hindrance).

HIND\_POL

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	13	17,3	19,4	19,4
	2,00	11	14,7	16,4	35,8
	3,00	25	33,3	37,3	73,1
	4,00	11	14,7	16,4	89,6
	5,00	7	9,3	10,4	100,0
	Total	67	89,3	100,0	
Missing	System	8	10,7		
Total		75	100,0		

3. Politicians are interested in PI (1 not interested – 5 highly interested)

**political interest**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	19	25,3	35,2	35,2
	2,00	18	24,0	33,3	68,5
	3,00	7	9,3	13,0	81,5
	4,00	8	10,7	14,8	96,3
	5,00	2	2,7	3,7	100,0
	Total	54	72,0	100,0	
Missing	System	21	28,0		
Total		75	100,0		

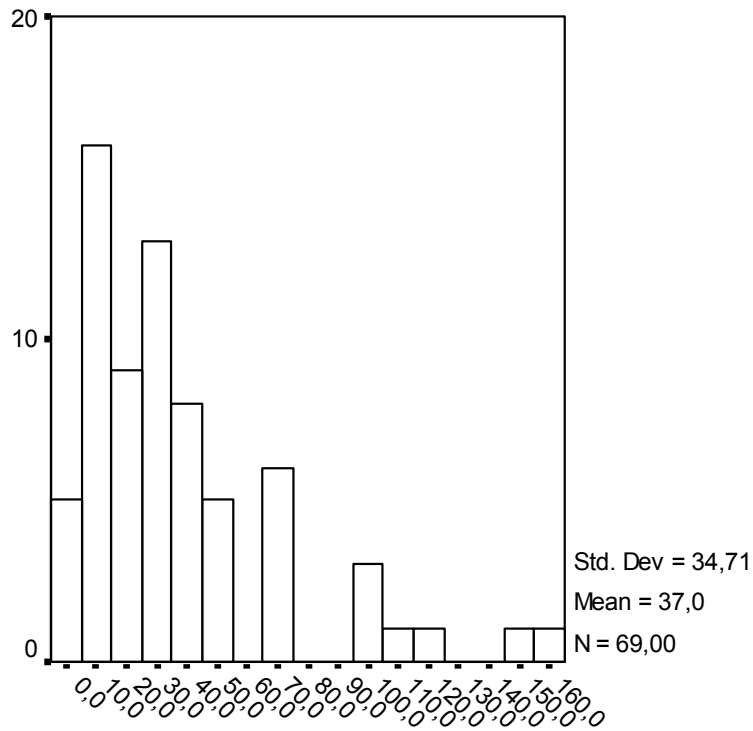
4. Time registration system (no, and increasingly detailed time segments (1-4))

**time registration**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	34	45,3	68,0	68,0
	1,00	2	2,7	4,0	72,0
	2,00	4	5,3	8,0	80,0
	3,00	2	2,7	4,0	84,0
	4,00	8	10,7	16,0	100,0
	Total	50	66,7	100,0	
Missing	System	25	33,3		
Total		75	100,0		

5. Full Time Equivalentents I

(excluding outliers with 538, 269, 226 and 225 fte)



fte tot

6. Lack of ICT resources (1 not a hindrance to 5 hindrance)

HIND\_ICT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	14	18,7	19,7	19,7
	2,00	20	26,7	28,2	47,9
	3,00	14	18,7	19,7	67,6
	4,00	14	18,7	19,7	87,3
	5,00	9	12,0	12,7	100,0
	Total	71	94,7	100,0	
Missing	System	4	5,3		
Total		75	100,0		

7. Lack of Financial resources (1 not a hindrance to 5 hindrance)

**HIND\_FIN**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	11	14,7	16,4	16,4
	2,00	14	18,7	20,9	37,3
	3,00	22	29,3	32,8	70,1
	4,00	18	24,0	26,9	97,0
	5,00	2	2,7	3,0	100,0
	Total	67	89,3	100,0	
Missing	System	8	10,7		
Total		75	100,0		

---

8. Lack of human resources (1 not a hindrance to 5 hindrance)

**HIND\_HR**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	5	6,7	7,1	7,1
	2,00	10	13,3	14,3	21,4
	3,00	18	24,0	25,7	47,1
	4,00	27	36,0	38,6	85,7
	5,00	10	13,3	14,3	100,0
	Total	70	93,3	100,0	
Missing	System	5	6,7		
Total		75	100,0		

---

9. Decoupling with objectives (1 not a hindrance to 5 hindrance)

**HIND\_GL**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	18	24,0	26,9	26,9
	2,00	22	29,3	32,8	59,7
	3,00	15	20,0	22,4	82,1
	4,00	10	13,3	14,9	97,0
	5,00	2	2,7	3,0	100,0
	Total	67	89,3	100,0	
Missing	System	8	10,7		
Total		75	100,0		

---



### 13.3. Descriptive statistics for chapter 7

**total demand**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	129	47,8	47,8	47,8
	1	82	30,4	30,4	78,1
	2	26	9,6	9,6	87,8
	3	17	6,3	6,3	94,1
	4	6	2,2	2,2	96,3
	5	5	1,9	1,9	98,1
	6	2	,7	,7	98,9
	7	1	,4	,4	99,3
	8	1	,4	,4	99,6
	10	1	,4	,4	100,0
	Total	270	100,0	100,0	

**demand (input)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	204	75,6	75,6	75,6
	1	47	17,4	17,4	93,0
	2	14	5,2	5,2	98,1
	3	4	1,5	1,5	99,6
	4	1	,4	,4	100,0
	Total	270	100,0	100,0	

**demand (intake)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	241	89,3	89,3	89,3
	1	25	9,3	9,3	98,5
	2	1	,4	,4	98,9
	3	2	,7	,7	99,6
	4	1	,4	,4	100,0
	Total	270	100,0	100,0	

**demand (output)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	216	80,0	80,0	80,0
	1	41	15,2	15,2	95,2
	2	9	3,3	3,3	98,5
	3	3	1,1	1,1	99,6
	4	1	,4	,4	100,0
	Total	270	100,0	100,0	

**demand (effect)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	235	87,0	87,0	87,0
	1	22	8,1	8,1	95,2
	2	8	3,0	3,0	98,1
	3	3	1,1	1,1	99,3
	4	1	,4	,4	99,6
	10	1	,4	,4	100,0
	Total	270	100,0	100,0	

**demand (context)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	262	97,0	97,0	97,0
	1	7	2,6	2,6	99,6
	2	1	,4	,4	100,0
	Total	270	100,0	100,0	

**total supply**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	145	53,7	53,7	53,7
	1	67	24,8	24,8	78,5
	2	24	8,9	8,9	87,4
	3	16	5,9	5,9	93,3
	4	9	3,3	3,3	96,7
	5	7	2,6	2,6	99,3
	7	1	,4	,4	99,6
	8	1	,4	,4	100,0
	Total	270	100,0	100,0	

**supply (input)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	213	78,9	78,9	78,9
	1	39	14,4	14,4	93,3
	2	13	4,8	4,8	98,1
	3	4	1,5	1,5	99,6
	5	1	,4	,4	100,0
	Total	270	100,0	100,0	

**supply (intake)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	238	88,1	88,1	88,1
	1	27	10,0	10,0	98,1
	2	2	,7	,7	98,9
	3	3	1,1	1,1	100,0
	Total	270	100,0	100,0	

**supply (output)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	216	80,0	80,0	80,0
	1	37	13,7	13,7	93,7
	2	11	4,1	4,1	97,8
	3	6	2,2	2,2	100,0
	Total	270	100,0	100,0	

**supply (effect)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	240	88,9	88,9	88,9
	1	21	7,8	7,8	96,7
	2	6	2,2	2,2	98,9
	3	1	,4	,4	99,3
	4	1	,4	,4	99,6
	7	1	,4	,4	100,0
	Total	270	100,0	100,0	

**supply (context)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	262	97,0	97,0	97,0
	1	7	2,6	2,6	99,6
	2	1	,4	,4	100,0
	Total	270	100,0	100,0	

**n° of direct links**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	175	64,8	64,8	64,8
	1	58	21,5	21,5	86,3
	2	20	7,4	7,4	93,7
	3	10	3,7	3,7	97,4
	4	4	1,5	1,5	98,9
	5	2	,7	,7	99,6
	7	1	,4	,4	100,0
	Total	270	100,0	100,0	

#### 13.4. Questionnaire

### **Questionnaire on performance measurement in the public sector**

January - February 2004

Prof. dr Geert Bouckaert

Wouter Van Dooren

Steunpunt Bestuurlijke Organisatie Vlaanderen

K.U.Leuven

## Section 1: identification of the section

1.1 What is the name of the section?

--

1.2 What is your function<sup>100</sup>?

A2A      A      B      C      D

1.3 In which phase of the policy cycle is your section mainly operating?

- predominantly policy preparation
- predominantly policy execution
- predominantly policy evaluation

(remarks)

1.4 In which policy sectors is your section for the moment (before BBB<sup>101</sup> reforms) operating? (multiple choices are possible)

<input type="radio"/>	Services of the Minister-president <sup>102</sup>	<input type="radio"/>	Culture, Sports and Media
<input type="radio"/>	Administrative Affairs <sup>103</sup>	<input type="radio"/>	Economy, Employment and Tourism
<input type="radio"/>	Finances and Budgeting	<input type="radio"/>	Environment and Nature
<input type="radio"/>	Foreign Affairs	<input type="radio"/>	Mobility
<input type="radio"/>	Science, Technology and Innovation	<input type="radio"/>	Spatial Planning, Housing Policy and immobile heritage
<input type="radio"/>	Education and Training	<input type="radio"/>	Agriculture and Fishery
<input type="radio"/>	Welfare and Public Health		

---

<sup>100</sup> The letters refer to the administrative grades in the Ministry of the Flemish Community: A2A is for a section manager; A is for professional (academically schooled staff); B C and D is for the rank and file staff.

<sup>101</sup> BBB stands for *Beter Bestuurlijk Beleid* (better policy of administration) - at the moment of the survey a pending reform of the Flemish public administration.

<sup>102</sup> Minister-president is the president of the regional government. Some general horizontal functions, such as administrative deregulation and Public Private Partnerships as well as study and statistical services are included in the services of the minister president.

<sup>103</sup> Administrative affairs includes amongst others administrative reform, local governments and provinces, human resources management and personnel affairs.

1.5 How many Full Time Equivalents are working in the section (on 01/01/2004) (effectively employed)?

Number of full time equivalents of level A                    ...  
Number of full time equivalents of level B and C                    ...  
Number of full time equivalents of level D                    ...

(remarks)

1.6 What is the total budget of the section (budget 2004)?

..... € (allocation in the budget) (excl. Personnel costs of the section)

(remarks)

1.7 Can you give an indication of the importance of the following costs in percentage of the budget of the section?

- operating budget ..... %  
- transfers (subsidies, dotations, ...) to families and companies ..... %  
- transfers (subsidies, dotaties, ...) to other organizations  
(e.g. hospitals, schools, municipalities, agencies, non profits, ...) ..... %

(remarks)

1.8 your section a regionally deconcentrated section<sup>104</sup>?

- yes
- no

---

<sup>104</sup> In administrative law, a distinction is made between deconcentration and decentralization. There are two kinds of decentralization in Flemish public administration. Regional decentralization refers to the transfer of competences to the local and sub-regional government. Functional decentralization refers to the transfer of competences to agencies within specific policy sectors. Deconcentration is different from decentralization because deconcentrated services remain under the hierarchical control of

## Section 2: identification of the measurement practices in the section.

Performance indicators tell something about the results of the organization. Indicators can refer to output of the organization as well as the effects of this output in society. The distinction between indicators on output and effect is crucial. The following table demonstrates this difference.

Output indicators are... Indicators that measure the quantity, quality and price of the services and products of the organization.	Effect indicators are ... Indicators that measure the effects in society if the products and services of the organization.
*what is leaving the organization...	...with which effect in the world outside
Examples The number of checked vehicles by the police The hours of television The number of training hours for unemployed	Examples Number of traffic casualties Viewing figures Number of participants that finds a job
In horizontal (supporting) sections, this distinction can be made too	
Examples number of wages paid in time the number of advance payments the number of payments which are controlled	Examples job satisfaction of the personnel fluid operations in the functional departments less fraude

2.1 2.1 Can you provide a typical output and a typical effect indicator??

(typical output indicator)  ...
---------------------------------------

(a typical outcome indicator)  ...
--

---

the central government. Yet, they are regionally dispersed in different locations in Flanders.



2.2 Some services and products are more easy to measure than others. What are the most important services/products of your section and to what extent do you think they are measurable (the question concerns measurability of the output, not the effect)?

Service/product	Assessment of the measurability of the output				
	Hard to measure			Easy to measure	
.....	1	2	3	4	5
.....	1	2	3	4	5
.....	1	2	3	4	5

(remarks)

2.3 Below are some fora for performance information about the output and the input of your section. To what extent does your section makes use if them?

*Are the indicators included in the policy plans and policy briefs of the ministers?*

- yes
- no

If yes, what is the number of indicators? ...

*Are the indicators of the section included in the accompanying documents to the budget?*

- yes
- no

If yes, what is the number of indicators? ...

*Does your section collect quantitative information on output and/or effect for answering MP's questions?*

- never
- once in a while
- approximately each month
- approximately each week
- approximately each day

*Can you indicate the **proportion** of parliamentary questions that requires quantitative information on output and effects on the total of parliamentary questions that the section answers?*

... % of the parliamentary questions that asks for quantitative information on output and/or effect?

*Does your section collect quantitative information about output and effect because of obligations in decrees<sup>105</sup>?*

- yes
- no

which decree?

*Does the section collect quantitative information on output and effect for use in the annual report?*

- yes
- no

*Does the section use quantitative information on output and effect of independent agencies to monitor the contracts with these independent agencies?*

- Yes
- No

---

<sup>105</sup> The highest law (in the material sense) of the regional parliament of Flanders are called decrees. A decree is in the hierarchy of laws at the same level of a federal law (in the formal sense).

Which institutions?

*Is quantitative information on output and effect collected for implementing management models?*

No management models (→ 2.25)

BSC (Balanced Scorecard)

EFQM (European Foundation for Quality Management)

ISO norms (International Standards Organization)

CAF (Common Assessment Framework)

Doelmatigheidsanalyse (efficiency analysis)

Other

*Is it an obligation from directorate-general<sup>106</sup>?*

- no
- a recommendation
- a formal obligation

*Is it an obligation of the secretariat-general?*

- no
- a recommendation
- a formal obligation

*j. Is quantitative information on output and/or effect collected to fulfil international reporting obligations?*

- yes
- no

---

<sup>106</sup> The Ministry of the Flemish Community has three hierarchical levels. The first level are the six departments which are headed by a secretary-general who has a staff - the secretariat general. The second level are the administrations which are headed by directors-general and his/her staff (directorate-general). The third level are the sections, which are headed by a section head. The respondents of the survey are the section heads.

If yes, which obligation (which institution)?

--

*k. Personal objectives are discussed during the annual planning conversation with the staff. For which percentage of the objectives is performance information being used?*

Planning conversation between director-general and the section manager	0% <input type="checkbox"/>	1%-25% <input type="checkbox"/>	26%-50% <input type="checkbox"/>	51%-75% <input type="checkbox"/>	76%-100% <input type="checkbox"/>
Planning conversation between the section manager and the staff on level A.	0% <input type="checkbox"/>	1%-25% <input type="checkbox"/>	26%-50% <input type="checkbox"/>	51%-75% <input type="checkbox"/>	76%-100% <input type="checkbox"/>
Planning conversation between the section manager and the staff on level B,C,D	0% <input type="checkbox"/>	1%-25% <input type="checkbox"/>	26%-50% <input type="checkbox"/>	51%-75% <input type="checkbox"/>	76%-100% <input type="checkbox"/>

(remarks)
-----------

2.4 The literature provides several potential hindrances for performance measurement. To what extent do you recognize these hindrances?

	Not a hindrance			A hindrance	
	1	2	3	4	5
ICT capacity is insufficient.					
There is no actual data to make meaningful indicators					
Indicators do not give a good idea of the measured reality					
We have insufficient financial resources to measure performance					
We have insufficient time (human resources) to measure performance.					
There is no political involvement					
The measurement effort is not supported by the personnel of the section					
The indicators are decoupled from the objectives of the section					
other: .....					
other .....					
other .....					

(remarks)
-----------

**WHEN YOUR SECTION IS IN NO WAY INVOLVED IN MEASUREMENT (CF. QUESTION 2.3), THEN THIS IS THE END OF THE QUESTIONNAIRE.**

2.5 When you look at the measurement practices of your section, to what extent does it cover the activities of your section?

- almost all the activities of your section
- a majority of the activities
- a minority of the activities
- almost no activities of the section

(remarks)

2.6 Does measurement concern predominantly output or effect-indicators?

- almost entirely effect-indicators
- predominantly effect-indicators
- as much effect as output indicators
- predominantly output indicators
- almost entirely output indicators

(remarks)

2.7 How often does your section discuss the performance of the section using indicators?

- seldom or never
- daily
- weekly
- monthly
- each trimester
- each semester
- yearly

Other frequency .....

(remarks)

2.8 Who participates in the discussion of the performance of your section (multiple choices possible)?

- section head
- responsible for a project or a team
- executive personnel
- director-general
- secretary-general
- representatives of political cabinets

other .....

(remarks)

2.9 To what extent does your section takes action based on performance information?

- never
- mostly not
- as many times as not
- mostly
- always

(remarks)

### Section 3: The production of performance information

3.1 Different groups can be involved in some way in the definition of the indicators. Which actors are involved in the definition of the indicators (output/effect) of your section?

	What is the involvement of the following actors in the definition of the indicators.					
	Not involved	Unimportant			Very important	
	0	1	2	3	4	5
The section	0	1	2	3	4	5
The executive personnel (cabinet of the) functional competent minister	0	1	2	3	4	5
(cabinet of the) horizontal competent minister	0	1	2	3	4	5
Other cabinets or meetings between cabinets.	0	1	2	3	4	5
The secretariat-general	0	1	2	3	4	5
het directorat-general	0	1	2	3	4	5
Other sections within the department	0	1	2	3	4	5
Other sections outside of the department	0	1	2	3	4	5
Internal audit	0	1	2	3	4	5
Agencies	0	1	2	3	4	5
Trade unions	0	1	2	3	4	5
Target group of the section (clients)	0	1	2	3	4	5
Members of parliament	0	1	2	3	4	5
international organizations	0	1	2	3	4	5
European institutions	0	1	2	3	4	5
Federal governments	0	1	2	3	4	5
Universities and polytechnics	0	1	2	3	4	5
Consultants	0	1	2	3	4	5
The third sector and interest groups	0	1	2	3	4	5
Other .....	0	1	2	3	4	5
Other .....	0	1	2	3	4	5
Other .....	0	1	2	3	4	5

(remarks)

3.2 Different data sources can be used for the indicators. Are the sources of performance indicators mainly internal or external?

- internal data collection (by the section itself)
- external data collection

(remarks)

3.3 What are the data sources of performance information (different choices are possible)? (give the data sources in order of importance where 1 is the most important source. If you do not use a particular source, you do not have to number it)

- ..... administrative registrations
- ..... extra registrations by the own staff
- ..... client surveys
- ..... self assessments
- ..... external observers
- ..... technical measurement
- ..... statistics institutions
- ..... international organizations
- ..... European organizations
- ..... federal public institutions
- ..... other Flemish public institutions
- ..... universities and polytechnics
- ..... other

(remarks)



3.4 An important aspect of performance information is the assessment whether the result is good or not. This assessment is often based on a comparison. Where is the comparison based on (multiple choices are possible)?

- comparisons in time (trend lines)
- comparisons between organizational divisions
- comparison with other organizations within the sector
- comparison with other organizations outside of the sector
- comparison with foreign organizations
- comparison with scientific standards
- comparison with negotiated or imposed norms
- comparison with organizations outside of the public sector
- other: .....

(remarks)

3.5 Does your organization have a system of time registration for the activities of the employees?

- yes, for more than 75% of the employees
- yes, for 50 to 75 % of the employees
- yes, for 25 to 50% of the employees
- yes, for less than 25% of the employees
- no (go to 3.7)

(Remarks)

3.6 What is the smallest registration unit?

- 15 min
- half a day
- fortnight
- half hour
- day
- month
- hour
- week
- other frequency: .....

(remarks)

3.7 Performance information of the section can be disseminated more or less freely within the section.  
Who has access to all the performance information?

- section manager
- section manager and some key person
- the whole section (go to 3.9)

remarks

3.8 What is the reason for not distributing some kinds of information within the whole organization?

Can you provide an example of a secured indicator?

3.9 Do you have explicit evaluations of the quality of the information?

- yes
- no (go to 3.11)

remarks)

3.10 What is the frequency of the evaluation of the quality of performance information?

..... times per year  
each ..... year

remarks

Does your section uses manuals on survey research and statistics from the administration for planning and statistics?

- almost never
- seldom
- often
- very often
- I do not know these manuals

remarks

3.11 When did the organization start with building a performance measurement system?

year .....

What was the concrete reason for collecting the indicators?

(remarks)

3.12 Statement: To what extent do you agree with the following statement?

“Political involvement was an important support for starting the measurement initiative”

Totally disagree		Totally agree		
1	2	3	4	5

(remarks)

3.13 Statement: To what extent do you agree with the following statement?

“At the moment, there is a strong political interest in the performance information of the section”

Totally disagree			Totally agree	
1	2	3	4	5

(remarks)
-----------

## Section 4: the use of performance information

4.1 Different actors may consume the performance information of the section. What are the most important consumers of the performance information of the section (multiple choices are possible)?

	Are the following actors important consumers of performance information?					
	Not involved	unimportant				Very important
	0	1	2	3	4	5
The section	0	1	2	3	4	5
The executive personnel (cabinet of the) functional competent minister	0	1	2	3	4	5
(cabinet of the) horizontal competent minister	0	1	2	3	4	5
Other cabinets or meetings between cabinets.	0	1	2	3	4	5
The secretariat-general	0	1	2	3	4	5
het directorat-general	0	1	2	3	4	5
Other sections within the department	0	1	2	3	4	5
Other sections outside of the department	0	1	2	3	4	5
Internal audit	0	1	2	3	4	5
Agencies	0	1	2	3	4	5
Trade unions	0	1	2	3	4	5
Target group of the section (clients)	0	1	2	3	4	5
Members of parliament	0	1	2	3	4	5
international organizations	0	1	2	3	4	5
European institutions	0	1	2	3	4	5
Federal governments	0	1	2	3	4	5
Universities and polytechnics	0	1	2	3	4	5
consultants	0	1	2	3	4	5
The third sector and interest groups	0	1	2	3	4	5
Other .....	0	1	2	3	4	5
Other .....	0	1	2	3	4	5
Other .....	0	1	2	3	4	5

remarks

4.2 Statement: To what extent do you agree with the following statement?

“The performance information of our section usually plays an important role political decision making.”

Totally disagree			Totally agree	
1	2	3	4	5

If you filled out 4 or 5, can you provide a recent example?

(remarks)

4.3 Performance information of organizations can sometimes play a central role in the public debate in the media and in politics. Examples are the viewing figures, or the number of refugees that are sent back. How often does performance information of your section play a role in the public debate?

- never (go to 4.5)
- once in a while
- multiple times a year
- multiple times a month
- very often

remarks

4.4 Which performance indicator was the last one debated in public? What was the theme?

4.5 Performance information can be used in different ways in your section. Can you indicate the intensity of use in your section?

Performance information is used in our section for...	Intensity					
	No use	Low intensity			High intensity	
Allocation of resources	0	1	2	3	4	5
Changing processes in the organization	0	1	2	3	4	5
Evaluating the section head	0	1	2	3	4	5
Evaluating the personnel	0	1	2	3	4	5
Managing teams	0	1	2	3	4	5
Human Resources Management	0	1	2	3	4	5
Changing policy measures	0	1	2	3	4	5
other .....	0	1	2	3	4	5
other .....	0	1	2	3	4	5
other .....	0	1	2	3	4	5

(remarks)

~ If you have additional remarks on the questionnaire or the subject, you can write them down below.

# Samenvatting

Prestatiemeting is een integraal onderdeel van de publieke sector. De laatste decennia nam de aandacht voor prestatie-informatie in de publieke sector duidelijk toe. Dit was onder meer te wijten aan de overheidshervormingen onder de vlag van het New Public Management. Het historisch overzicht in deze studie toont echter aan dat de roots van prestatiemeting heel wat verder terug gaan. Het meten van de prestaties van de publieke sector is iets van alle tijden. Wel is het zo dat prestatiemeting doorheen de tijd in steeds sterkere mate alle sectoren van de overheid infiltreerde.

Deze studie streeft naar een beter begrip van prestatiemeting in de publieke sector. De literatuurstudie toonde aan dat prestatiemeting en prestatie-management vaak op een ondoorzichtige wijze worden geoperationaliseerd, en vaak normatief worden ingevuld. Deze studie maakte een duidelijk onderscheid tussen vraag en aanbod van prestatie-informatie. Met dit onderscheid als uitgangspunt, werden vier onderzoeksvragen onderzocht. De methodologie was een combinatie van survey, interviews en documentenanalyse.

1. Om welke redenen en onder welke condities zullen organisaties overgaan tot meten? We formuleerden zes hypothesen waarom organisaties meten. We maakten een onderscheid tussen de adoptie van prestatiemeting enerzijds en de implementatie van prestatiemeting anderzijds.

2. Komen administratief aanbod en politieke vraag overeen, en is er een verschil tussen beleidssectoren? Aan de hand van parlementaire vragen gingen we na wat de kwantiteit en de kwaliteit van de match tussen vraag en aanbod was. We vonden dat metingen in meer dan de helft van de vragen terugkwam. Er waren significante verschillen tussen sectoren op het vlak van het aantal en de focus van de indicatoren.

3. Hoe kan het gebruik van prestatie-informatie het design van een meetsysteem bepalen? We maken een onderscheid tussen drie categorieën van gebruik – onderzoek en leren, intern management en accountability. Deze gebruikswijze beperken in toenemende mate de vrijheidsgraden van organisatie. Dit heeft repercussies voor alle stappen in het productieproces.

4. Wat zijn de effecten van prestatiemeting op organisaties? Onderzoekers hebben reeds heel wat effecten geïdentificeerd. Wat vaak ontbrak was een empirische toetsing van deze effecten. In deze studie maakten we een onderscheid tussen de meeteffecten en gedragseffecten. We maakten een inschatting van de kans op verschillende gedragseffecten en verklaarden de effecten door middel van het gebruik van prestatie-informatie.



## Résumé

La mesure de la performance est une partie intégrale du secteur public. Les dernières années, l'attention pour la mesure de la performance a augmenté. L'impulse était entre autres les réformes dans le cadre de la Nouvelle Gestion Publique. Le chapitre historique a néanmoins démontré que les origines de la mesure de la performance sont plus vieilles. On a toujours mesuré la performance du secteur publique, mais à travers les ages, la mesure de la performance s'est intégrée dans tous les secteurs du gouvernement.

Notre étude aspire une meilleure compréhension de la mesure de la performance dans le secteur public. L'étude de la littérature a démontré que la mesure et la gestion de la performance sont opérationnalisés sur une manière troublée et normative. Notre étude fait la distinction transparente entre l'offre et la demande pour l'information. Avec cette distinction comme point de départ, quatre questions de recherche ont été étudiées. La méthodologie est une combinaison de survey, des interviews, et l'analyse des documents.

1. Quelles sont les conditions et les motivations pour lesquelles organisations commencent à mesurer? Nous avons formulé six hypothèses. Nous faisons une distinction entre l'adoption et l'implémentation de la mesure de performance.

2. Est-ce qu'il y a une correspondance entre l'offre administrative et la demande politique, et est-ce qu'il y a une différence entre les secteurs? On a analysé la quantité et la qualité de la correspondance entre offre et demande au moyen de l'analyse des questions parlementaires. On a retrouvé que des mesures dans la moitié des questions. Il y avait des différences significatives entre les secteurs sur le plan de la quantité et sur le plan du sujet des indicateurs.

3. Comment l'utilisation peut déterminer le design d'un système de mesurer. Nous distinguons trois catégories d'utilisation – la recherche et l'apprentissage, la gestion interne, et l'accountability. Ces trois modes d'utilisation limitent différemment les degrés de liberté de l'organisation. La différence a des répercussions sur toutes les phases du processus de production.

4. Quels sont les effets de la mesure de la performance? Des chercheurs ont déjà identifié une liste des effets. Ce qui manque ces études est la teste empirique. Notre étude fait une distinction entre des effets sur la système de mesurer et des effets sur la comportement. Nous avons fait une estimation de la propension de différents effets de comportement et nous avons expliqué la différence par rapport à l'utilisation de l'information.

# DOCTORATEN IN DE SOCIALE WETENSCHAPPEN

## I. REEKS VAN DOCTORATEN IN DE SOCIALE WETENSCHAPPEN<sup>(1)</sup>

1. CLAEYS, U., *De sociale mobiliteit van de universitair afgestudeerden te Leuven. Het universitair onderwijs als mobiliteitskanaal*, 1971, 2 delen 398 blz.
2. VANHESTE, G., *Literatuur en revolutie*, 1971, 2 delen, 500 blz.
3. DELANGHE, L., *Differentiële sterfte in België. Een sociaal-demografische analyse*, 1971, 3 delen, 773 blz.
4. BEGHIN, P., *Geleide verandering in een Afrikaanse samenleving. De Bushi in de koloniale periode*, 1971, 316 blz.
5. BENOIT, A., *Changing the education system. A Colombian case-study*, 1972, 382 blz.
6. DEFEVER, M., *De huisartssituatie in België*, 1972, 374 blz.
7. LAUWERS, J., *Kritische studie van de secularisatietheorieën in de sociologie*, 1972, 364 blz.
8. GHOOS, A., *Sociologisch onderzoek naar de gevolgen van industrialisering in een rekonversiegebied*, 1972, 256 blz. + bijlagen.
9. SLEDESENS, G., *Mariage et vie conjugale du moniteur rwandais. Enquête sociologique par interview dirigée parmi les moniteurs mariés rwandais*, 1972, 2 delen, 549 blz.
10. TSAI, C., *La chambre de commerce internationale. Un groupe de pression international. Son action et son rôle dans l'élaboration, la conclusion et l'application des conventions internationales établies au sein des organisations intergouvernementales à vocation mondiale (1945-1969)*, 1972, 442 blz.

---

<sup>(1)</sup> EEN EERSTE SERIE DOCTORATEN VORMT DE REEKS VAN DE SCHOOL VOOR POLITIEKE EN SOCIALE WETENSCHAPPEN (NRS. 1 TOT EN MET 185). DE INTEGRALE LIJST KAN WORDEN GEVONDEN IN NADIEN GEPUBLICEERDE DOCTORATEN, ZOALS G. DOOGHE, "DE STRUCTUUR VAN HET GEZIN EN DE SOCIALE RELATIES VAN DE BEJAARDEN". ANTWERPEN, DE NEDERLANDSE BOEKHANDEL, 1970, 290 BLZ.

EEN TWEEDE SERIE DOCTORATEN IS VERMELD IN DE "NIEUWE REEKS VAN DE FACULTEIT DER ECONOMISCHE EN SOCIALE WETENSCHAPPEN". DE INTEGRALE LIJST KAN WORDEN GEVONDEN IN O.M. M. PEETERS, "GODSDIENST EN TOLERANTIE IN HET SOCIALISTISCH DENKEN". EEN HISTORISCH-DOCTRINAIRE STUDIE, 1970, 2 DELEN, 568 BLZ.

11. DEPRE, R., *De topambtenaren van de ministeries in België. Een bestuurssociologisch onderzoek*, 1973, 2 delen, 423 blz. + bijlagen.
  
12. VAN DER BIESEN, W., *De verkiezingspropaganda in de democratische maatschappij. Een literatuurkritische studie en een inhoudsanalyse van de verkiezingscampagne van 1958 in de katholieke pers en in de propagandapublicaties van de C.V.P.*, 1973, 434 blz.
  
13. BANGO, J., *Changements dans les communautés villageoises de l'Europe de l'Est. Exemple : la Hongarie*, 1973, 434 blz.
  
14. VAN PELT, H., *De omroep in revisie. Structurering en ontwikkelingsmogelijkheden van het radio- en televisiebestel in Nederland en België. Een vergelijkende studie*, Leuven, Acco, 1973, 398 blz.
  
15. MARTENS, A., *25 jaar wegwerparbeiders. Het Belgisch immigratiebeleid na 1945*, 1973, 319 blz.
  
16. BILLET, M., *Het verenigingsleven in Vlaanderen. Een sociologische typologieformulering en hypothesetoetsing*, 1973, 695 blz. + bijlagen.
  
17. BRUYNOOGHE, R., *De sociale structurering van de gezinsverplegingssituatie vanuit kostgezinnen en patiënten*, 1973, 205 blz. + bijlagen.
  
18. BUNDERVOET, J., *Het doorstromingsprobleem in de hedendaagse vakbeweging. Kritische literatuurstudie en verkennend onderzoek in de Belgische vakbonden*, 1973, 420 blz. + bijlagen.
  
19. GEVERS, P., *Ondernemingsraden, randverschijnselen in de Belgische industriële democratiseringsbeweging. Een sociologische studie*, 1973, 314 blz.
  
20. MBELA, H., *L'intégration de l'éducation permanente dans les objectifs socio-économiques de développement. Analyse de quelques politiques éducationnelles en vue du développement du milieu rural traditionnel en Afrique noire francophone*, 1974, 250 blz.
  
21. CROLLEN, L., *Small powers in international systems*, 1974, 250 blz.
  
22. VAN HASSEL, H., *Het ministrieel kabinet. Peilen naar een sociologische duiding*, 1974, 460 blz. + bijlagen.
  
23. MARCK, P., *Public relations voor de landbouw in de Europese Economische Gemeenschap*, 1974, 384 blz.
  
24. LAMBRECHTS, E., *Vrouwenarbeid in België. Een analyse van het tewerkstellingsbeleid inzake vrouwelijke arbeidskrachten sinds 1930*, 1975, 260 blz.

25. LEMMEN, M.H.W., *Rationaliteit bij Max Weber. Een godsdienstsociologische studie*, 1975, 2 delen, 354 blz.
26. BOON, G., *Ontstaan, ontwikkeling en werking van de radio-omroep in Zaïre tijdens het Belgisch Koloniale Bewind (1937-1960)*, 1975, 2 delen, 617 blz.
27. WUYTS, H., *De participatie van de burgers in de besluitvorming op het gebied van de gemeentelijke plannen van aanleg. Analyse toegespitst op het Nederlandstalige deel van België*, 1975, 200 blz. + bijlage.
28. VERRIEST, F., *Joris Helleputte en het corporatisme*, 1975, 2 delen, 404 blz.
29. DELMARTINO, F., *Schaalvergroting en bestuurskracht. Een beleidsanalytische benadering van de herstructurering van de lokale besturen*, 1975, 3 delen, 433 blz. + bijlagen.
30. BILLIET, J., *Secularisering en verzuiling in het Belgisch onderwijs*, 1975, 3 delen, 433 blz. + bijlagen.
31. DEVISCH, R., *L'institution rituelle Khita chez les Yaka au Kwaango du Nord. Une analyse sémiologique*, 1976, 3 volumes.
32. LAMMERTYN, F., *Arbeidsbemiddeling en werkloosheid. Een sociologische verkenning van het optreden van de diensten voor openbare arbeidsbemiddeling van de R.V.A.*, 1976, 406 blz.
33. GOVAERTS, F., *Zwitserland en de E.E.G. Een case-study inzake Europese integratie*, 1976, 337 blz.
34. JACOBS, T., *Het uit de echt scheiden. Een typologiserend onderzoek, aan de hand van de analyse van rechtsplegingsdossiers in echtscheiding*. 1976, 333 blz. + bijlage.
35. KIM DAI WON, *Au delà de l'institutionnalisation des rapports professionnels. Analyse du mouvement spontané ouvrier belge*. 1977, 282 blz.
36. COLSON, F., *Sociale indicatoren van enkele aspecten van bevolkingsgroei*. 1977, 341 blz. + bijlagen.
37. BAECK, A., *Het professionaliseringsproces van de Nederlandse huisarts*. 1978, 721 blz. + bibliografie.
38. VLOEBERGHS, D., *Feedback, communicatie en organisatie. Onderzoek naar de betekenis en de toepassing van het begrip "feedback" in de communicatiewetenschap en de organisatietheorieën*. 1978, 326 blz.
39. DIERICKX, G., *De ideologische factor in de Belgische politieke besluitvorming*. 1978, 609 blz. + bijvoegsels.
40. VAN DE KERCKHOVE, J., *Sociologie. Maatschappelijke relevantie en arbeidersemancipatie*. 1978, 551 blz.

41. DE MEYER A., *De populaire muziekindustrie. Een terreinverkennde studie.* 1979, 578 blz.
42. UDDIN, M., *Some Social Factors influencing Age at Death in the situation of Bangladesh.* 1979, 316 blz. + bijlagen.
43. MEULEMANS, E., *De ethische problematiek van het lijden aan het leven en aan het samen-leven in het oeuvre van Albert Camus. De mogelijke levensstijlen van luciditeit, menselijkheid en solidariteit.* 1979, 413 blz.
44. HUYPENS, J., *De plaatselijke nieuwsfabriek. Regionaal nieuws. Analyse van inhoud en structuur in de krant.* 494 blz.
45. CEULEMANS, M.J., *Women and Mass Media: a feminist perspective. A review of the research to date the image and status of women in American mass media.* 1980, 541 blz. + bijlagen.
46. VANDEKERCKHOVE, L., *Gemaakt van asse. Een sociologische studie van de westerse somatische cultuur.* 1980, 383 blz.
47. MIN, J.K., *Political Development in Korea, 1945-1972.* 1980, 2 delen, 466 blz.
48. MASUI, M., *Ongehuwd moeder. Sociologische analyse van een wordingsproces.* 1980, 257 blz.
49. LEDOUX, M., *Op zoek naar de rest ...; Genealogische lezing van het psychiatrisch discours.* 1981, 511 blz.
50. VEYS, D., *De generatie-sterftetafels in België.* 1981, 3 delen, 326 blz. + bijlagen.
51. TACQ, J., *Kausaliteit in sociologisch onderzoek. Een beoordeling van de zgn. 'causal modeling'-technieken in het licht van verschillende wijsgerige opvattingen over kausaliteit.* 1981, 337 blz.
52. NKUNDABAGENZI, F., *Le système politique et son environnement. Contribution à l'étude de leur interaction à partir du cas des pays est-africains : le Kenya et la Tanzanie.* 1981, 348 blz.
53. GOOSSENS, L., *Het sociaal huisvestingsbeleid in België. Een historisch-sociologische analyse van de maatschappelijke probleembehandeling op het gebied van het wonen.* 1982, 3 delen.
54. SCHEPERS, R., *De opkomst van het Belgisch medisch beroep. De evolutie van de wetgeving en de beroepsorganisatie in de 19de eeuw.* 1983, 553 blz.
55. VANSTEENKISTE, J., *Bejaardzijn als maatschappelijk gebeuren.* 1983, 166 blz.
56. MATTHIJS, K., *Zelfmoord en zelfmoordpoging.* 1983, 3 delen, 464 blz.

57. CHUNG-WON, Choue, *Peaceful Unification of Korea. Towards Korean Integration.* 1984, 338 blz.
58. PEETERS, R., *Ziekte en gezondheid bij Marokkaanse immigranten.* 1983, 349 blz.
59. HESLING, W., *Retorica en film. Een onderzoek naar de structuur en functie van klassieke overtuigingsstrategieën in fictionele, audiovisuele teksten.* 1985, 515 blz.
60. WELLEN, J., *Van probleem tot hulpverlening. Een exploratie van de betrekkingen tussen huisartsen en ambulante geestelijke gezondheidszorg in Vlaanderen.* 1984, 476 blz.
61. LOOSVELDT, G., *De effecten van een interviewtraining op de kwaliteit van gegevens bekomen via het survey-interview.* 1985, 311 blz. + bijlagen.
62. FOETS, M., *Ziekte en gezondheidsgedrag : de ontwikkeling van de sociologische theorievorming en van het sociologisch onderzoek.* 1985, 339 blz.
63. BRANCKAERTS, J., *Zelfhulporganisaties. Literatuuranalyse en explorerend onderzoek in Vlaanderen.* 1985.
64. DE GROOFF, D., *De elektronische krant. Een onderzoek naar de mogelijkheden van nieuwsverspreiding via elektronische tekstmedia en naar de mogelijke gevolgen daarvan voor de krant als bedrijf en als massamedium.* 1986, 568 blz.
65. VERMEULEN, D., *De maatschappelijke beheersingsprocessen inzake de sociaal-culturele sector in Vlaanderen. Een sociologische studie van de "verzuiling", de professionalisering en het overheidsbeleid.* 1983, 447 blz.
66. OTSHOMANPITA, Aloki, *Administration locale et développement au Zaïre. Critiques et perspectives de l'organisation politico-administrative à partir du cas de la zone de Lodja.* 1988, 507 blz.
67. SERVAES, J., *Communicatie en ontwikkeling. Een verkennende literatuurstudie naar de mogelijkheden van een communicatiebeleid voor ontwikkelingslanden.* 1987, 364 blz.
68. HELLEMANS, G., *Verzuiling. Een historische en vergelijkende analyse.* 1989, 302 blz.

## **II. NIEUWE REEKS VAN DOCTORATEN IN DE SOCIALE WETENSCHAPPEN**

1. LIU BOLONG, *Western Europe - China. A comparative analysis of the foreign policies of the European Community, Great Britain and Belgium towards China (1970-1986).* Leuven, Departement Politieke Wetenschappen, 1988, 335 blz.
2. EERDEKENS, J., *Chronische ziekte en rolverandering. Een sociologisch onderzoek bij M.S.-patiënten.* Leuven, Acco, 1989, 164 blz. + bijlagen.

3. HOUBEN, P., *Formele beslissingsmodellen en speltheorie met toepassingen en onderzoek naar activiteiten en uitgaven van lokale welzijnsinstellingen en coalities*. Leuven, Departement Sociologie, 1988, 631 blz. (5 delen).
4. HOOGHE, L., *Separatisme. Conflict tussen twee projecten voor natievorming. Een onderzoek op basis van drie succesvolle separatismen*. Leuven, Departement Politieke Wetenschappen, 1989, 451 blz. + bijlagen.
5. SWYNGEDOUW, M., *De keuze van de kiezer. Naar een verbetering van de schattingen van verschuivingen en partijvoorkeur bij opeenvolgende verkiezingen en peilingen*. Leuven, Sociologisch Onderzoeksinstituut, 1989, 333 blz.
6. BOUCKAERT, G., *Productiviteit in de overheid*. Leuven, Vervolmakingscentrum voor Overheidsbeleid en Bestuur, 1990, 394 blz.
7. RUEBENS, M., *Sociologie van het alledaagse leven*. Leuven, Acco, 1990, 266 blz.
8. HONDEGHEM, A., *De loopbaan van de ambtenaar. Tussen droom en werkelijkheid*. Leuven, Vervolmakingscentrum voor Overheidsbeleid en Bestuur, 1990, 498 blz. + bijlage.
9. WINNUBST, M., *Wetenschapspopularisering in Vlaanderen. Profiel, zelfbeeld en werkwijze van de Vlaamse wetenschapsjournalist*. Leuven, Departement Communicatiewetenschap, 1990.
10. LAERMANS, R., *In de greep van de "moderne tijd". Modernisering en verzuiling, individualisering en het naoorlogse publieke discours van de ACW-vormingsorganisaties : een proeve tot cultuursociologische duiding*. Leuven, Garant, 1992.
11. LUYTEN, D., *OCMW en Armeazorg. Een sociologische studie van de sociale grenzen van het recht op bijstand*. Leuven, S.O.I. Departement Sociologie, 1993, 487 blz.
12. VAN DONINCK, B., *De landbouwcoöperatie in Zimbabwe. Bouwsteen van een nieuwe samenleving ?* Grimbergen, vzw Belgium-Zimbabwe Friendship Association, 1993. 331 blz.
13. OPDEBEECK, S., *Afhankelijkheid en het beëindigen van partnergeweld*. Leuven, Garant, 1993. 299 blz. + bijlagen.
14. DELHAYE, C., *Mode geleefd en gedragen*. Leuven, Acco, 1993, 228 blz.
15. MADDENS, B., *Kiesgedrag en partijstrategie*. Leuven, Departement Politieke Wetenschappen, Afdeling Politologie, K.U.Leuven, 1994, 453 blz.
16. DE WIT, H., *Cijfers en hun achterliggende realiteit. De MTMM-kwaliteitsparameters op hun kwaliteit onderzocht*. Leuven, Departement Sociologie, K.U.Leuven, 1994, 241 blz.

17. DEVELTERE, P., *Co-operation and development with special reference to the experience of the Commonwealth Carribean*. Leuven, Acco, 1994, 241 blz.
  
18. WALGRAVE, S., *Tussen loyaliteit en selectiviteit. Een sociologisch onderzoek naar de ambivalente verhouding tussen nieuwe sociale bewegingen en groene partij in Vlaanderen*. Leuven, Garant, 1994, 361 blz.
  
19. CASIER, T., *Over oude en nieuwe mythen. Ideologische achtergronden en repercussies van de politieke omwentelingen in Centraal- en Oost-Europa sinds 1985*. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 1994, 365 blz.
  
20. DE RYNCK, F., *Streekontwikkeling in Vlaanderen. Besturingsverhoudingen en beleidsnetwerken in bovenlokale ruimtes*. Leuven, Departement Politieke Wetenschappen, Afdeling Bestuurswetenschap, K.U.Leuven, 1995, 432 blz.
  
21. DEVOS, G., *De flexibilisering van het secundair onderwijs in Vlaanderen. Een organisatie-sociologische studie van macht en institutionalisering*. Leuven, Acco, 1995, 447 blz.
  
22. VAN TRIER, W., *Everyone A King? An investigation into the meaning and significance of the debate on basic incomes with special references to three episodes from the British inter-War experience*. Leuven, Departement Sociologie, K.U.Leuven, 1995, vi+501 blz.
  
23. SELS, L., *De overheid viert de teugels. De effecten op organisatie en personeelsbeleid in de autonome overheidsbedrijven*. Leuven, Acco, 1995, 454 blz.
  
24. HONG, K.J., *The C.S.C.E. Security Regime Formation: From Helsinki to Budapest*. Leuven, Acco, 1996, 350 blz.
  
25. RAMEZANZADEH, A., *Internal and international dynamics of ethnic conflict. The Case of Iran*. Leuven, Acco, 1996, 273 blz.
  
26. HUYSMANS, J., *Making/Unmaking European Disorder. Meta-Theoretical, Theoretical and Empirical Questions of Military Stability after the Cold War*. Leuven, Acco, 1996, 250 blz.
  
27. VAN DEN BULCK J., *Kijkbuis kennis. De rol van televisie in de sociale en cognitieve constructie van de realiteit*. Leuven, Acco, 1996, 242 blz.
  
28. JEMADU Aleksius, *Sustainable Forest Management in the Context of Multi-level and Multi-actor Policy Processes*. Leuven, Departement Politieke Wetenschappen, Afdeling Bestuur en Overheidsmanagement, K.U.Leuven, 1996, 310 blz.
  
29. HENDRAWAN Sanerya, *Reform and Modernization of State Enterprises. The Case of Indonesia*. Leuven, Departement Politieke Wetenschappen, Afdeling Bestuur en Overheidsmanagement, K.U.Leuven, 1996, 372 blz.



30. MUIJS Roland Daniël, *Self, School and Media: A Longitudinal Study of Media Use, Self-Concept, School Achievement and Peer Relations among Primary School Children*. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 1997, 316 blz.
31. WAEGE Hans, *Vertogen over de relatie tussen individu en gemeenschap*. Leuven, Acco, 1997, 382 blz.
32. FIERS Stefaan, *Partijvoorzitters in België of 'Le parti, c'est moi'?* Leuven, Acco, 1998, 419 blz.
33. SAMOY Erik, *Ongeschied of ongewenst? Een halve eeuw arbeidsmarktbeleid voor gehandicapten*. Leuven, Departement Sociologie, K.U.Leuven, 1998, 640 blz.
34. KEUKELEIRE Stephan, *Het Gemeenschappelijk Buitenlands en Veiligheidsbeleid (GBVB): het buitenlands beleid van de Europese Unie op een dwaalspoor*. Leuven, Departement Politieke Wetenschappen, Afdeling Internationale Betrekkingen, K.U.Leuven, 1998, 452 blz.
35. VERLINDEN Ann, *Het ongewone alledaagse: over zwarte katten, horoscopen, miraculeuze genezingen en andere geloofselementen en praktijken. Een sociologie van het zogenaamde bijgeloof*. Leuven, Departement Sociologie, K.U.Leuven, 1999, 387 blz. + bijlagen.
36. CARTON Ann, *Een interviewernetwerk: uitwerking van een evaluatieprocedure voor interviewers*. Leuven, Departement Sociologie, 1999, 379 blz. + bijlagen.
37. WANG Wan-Li, *Understanding Taiwan-EU Relations: An Analysis of the Years from 1958 to 1998*. Leuven, Departement Politieke Wetenschappen, Afdeling Internationale Betrekkingen, K.U.Leuven, 1999, 326 blz. + bijlagen.
38. WALRAVE Michel, *Direct Marketing en Privacy. De verhouding tussen direct marketingscommunicatie en de bescherming van de informationele en de relationele privacy van consumenten*. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 1999, 480 blz. + bijlagen.
39. KOCHUYT Thierry, *Over een ondercultuur. Een cultuursociologische studie naar de relatieve deprivatie van arme gezinnen*. Leuven, Departement Sociologie, K.U.Leuven, 1999, 386 blz. + bijlagen.
40. WETS Johan, *Waarom onderweg? Een analyse van de oorzaken van grootschalige migratie- en vluchtelingenstromen*. Leuven, Departement Politieke Wetenschappen, Afdeling Internationale Betrekkingen, K.U.Leuven, 1999, 321 blz. + bijlagen.
41. VAN HOOTEGEM Geert, *De draaglijke traagheid van het management. Productie- en Personeelsbeleid in de industrie*. Leuven, Departement Sociologie, K.U.Leuven, 1999, 471 blz. + bijlagen.
42. VANDEBOSCH Heidi, *Een geboeid publiek? Het gebruik van massamedia door gedetineerden*. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 1999, 375 blz. + bijlagen.
43. VAN HOVE Hildegard, *De weg naar binnen. Spiritualiteit en zelfontplooiing*. Leuven, Departement Sociologie, K.U.Leuven, 2000, 369 blz. + bijlagen.

44. HUYS Rik, *Uit de band? De structuur van arbeidsverdeling in de Belgische autoassemblagebedrijven*. Leuven, Departement Sociologie, K.U.Leuven, 2000, 464 blz. + bijlagen.
  
45. VAN RUYSEVELDT Joris, *Het belang van overleg. Voorwaarden voor macroresponsieve CAO-onderhandelingen in de marktsector*. Leuven, Departement Sociologie, K.U.Leuven, 2000, 349 blz. + bijlagen.
  
46. DEPAUW Sam, *Cohesie in de parlementsfracties van de regeringsmeerderheid. Een vergelijkend onderzoek in België, Frankrijk en het Verenigd Koninkrijk (1987-97)*. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2000, 510 blz. + bijlagen.
  
47. BEYERS Jan, *Het maatschappelijk draagvlak van het Europees beleid en het einde van de permissieve consensus. Een empirisch onderzoek over politiek handelen in een meerlagig politiek stelsel*. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2000, 269 blz. + bijlagen.
  
48. VAN DEN BULCK Hilde, *De rol van de publieke omroep in het project van de moderniteit. Een analyse van de bijdrage van de Vlaamse publieke televisie tot de creatie van een nationale cultuur en identiteit (1953-1973)*. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2000, 329 blz. + bijlagen.
  
49. STEEN Trui, *Krachtlijnen voor een nieuw personeelsbeleid in de Vlaamse gemeenten. Een studie naar de sturing en implementatie van veranderingsprocessen bij de overheid*. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2000, 340 blz. + bijlagen.
  
50. PICKERY Jan, *Applications of Multilevel Analysis in Survey Data Quality Research. Random Coefficient Models for Respondent and Interviewer Effects*. Leuven, Departement Sociologie, K.U.Leuven, 2000, 200 blz. + bijlagen.
  
51. DECLERCQ Aniana (Anja), *De complexe zoektocht tussen orde en chaos. Een sociologische studie naar de differentiatie in de institutionele zorgregimes voor dementerende ouderen*. Leuven, Departement Sociologie, K.U.Leuven, 2000, 260 blz. + bijlagen.
  
52. VERSCHRAEGEN Gert, *De maatschappij zonder eigenschappen. Systeemtheorie, sociale differentiatie en moraal*. Leuven, Departement Sociologie, K.U.Leuven, 2000, 256 blz. + bijlagen.
  
53. DWIKARDANA Saptia, *The Political Economy of Development and Industrial Relations in Indonesia under the New Order Government*. Leuven, Departement Sociologie, K.U.Leuven, 2001, 315 blz. + bijlagen.
  
54. SAUER Tom, *Nuclear Inertia. US Nuclear Weapons Policy after the Cold War (1990-2000)*. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2001, 358 blz. + bijlagen.
  
55. HAJNAL Istvan, *Classificatie in de sociale wetenschappen. Een evaluatie van de nauwkeurigheid van een aantal clusteranalysemethoden door middel van simulaties*. Leuven, Departement Sociologie, K.U.Leuven, 2001, 340 blz. + bijlagen.

56. VAN MEERBEECK Anne, Het doopsel: een familieritueel. Een sociologische analyse van de betekenissen van dopen in Vlaanderen. Leuven, Departement Sociologie, K.U.Leuven, 2001, 338 blz. + bijlagen.
57. DE PRINS Peggy, Zorgen om zorg(arbeid). Een vergelijkend onderzoek naar oorzaken van stress en maatzorg in Vlaamse rusthuizen. Leuven, Departement Sociologie, K.U.Leuven, 2001, 363 blz. + bijlagen.
58. VAN BAVEL Jan, Demografische reproductie en sociale evolutie: geboortebeperving in Leuven 1840-1910. Leuven, Departement Sociologie, K.U.Leuven, 2001, 362 blz. + bijlagen.
59. PRINSLOO Riana, Subnationalism in a Cleavaged Society with Reference to the Flemish Movement since 1945. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2001, 265 blz. + bijlagen.
60. DE LA HAYE Jos, Missed Opportunities in Conflict Management. The Case of Bosnia-Herzegovina (1987-1996). Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2001, 283 blz. + bijlagen.
61. ROMMEL Ward, Heeft de sociologie nood aan Darwin? Op zoek naar de verhouding tussen evolutiepsychologie en sociologie. Leuven, Departement Sociologie, K.U.Leuven, 2002, 287 blz. + bijlagen.
62. VERVLIET Chris, Vergelijking tussen Duits en Belgisch federalisme, ter toetsing van een neofunctionalistisch verklaringsmodel voor bevoegdheidsverschuivingen tussen nationale en subnationale overheden: een analyse in het economisch beleidsdomein. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 265 blz. + bijlagen.
63. DHOEST Alexander, De verbeelde gemeenschap: Vlaamse tv-fictie en de constructie van een nationale identiteit. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2002, 384 blz. + bijlagen.
64. VAN REETH Wouter, The Bearable Lightness of Budgeting. The Uneven Implementation of Performance Oriented Budget Reform Across Agencies. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 380 blz. + bijlagen.
65. CAMBRÉ Bart, De relatie tussen religiositeit en ethnocentrisme. Een contextuele benadering met cross-culturele data. Leuven, Departement Sociologie, K.U.Leuven, 2002, 257 blz. + bijlagen.
66. SCHEERS Joris, Koffie en het aroma van de stad. Tropische (re-)productiestructuren in ruimtelijk perspectief. Casus centrale kustvlakte van Ecuador. Leuven, Departement Sociologie, K.U.Leuven, 2002, 294 blz. + bijlagen.
67. VAN ROMPAEY Veerle, Media on / Family off? An integrated quantitative and qualitative investigation into the implications of Information and Communication Technologies (ICT) for family life. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2002, 232 blz. + bijlagen.
68. VERMEERSCH Peter, Roma and the Politics of Ethnicity in Central Europe. A Comparative Study of Ethnic Minority Mobilisation in the Czech Republic, Hungary and Slovakia in the 1990s. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 317 blz. + bijlagen.

69. GIELEN Pascal, Pleidooi voor een symmetrische kunstsociologie. Een sociologische analyse van artistieke selectieprocessen in de sectoren van de hedendaagse dans en de beeldende kunst in Vlaanderen. Leuven, Departement Sociologie, K.U.Leuven, 2002, 355 blz. + bijlagen.
  
70. VERHOEST Koen, Resultaatgericht verzelfstandigen. Een analyse vanuit een verruimd principaal-agent perspectief. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2002, 352 blz. + bijlagen.
  
71. LEFÈVRE Pascal, Willy Vandersteens Suske en Wiske in de krant (1945-1971). Een theoretisch kader voor een vormelijke analyse van strips. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2003, 186 blz. (A3) + bijlagen.
  
72. WELKENHUYSEN-GYBELS Jerry, The Detection of Differential Item Functioning in Likert Score Items. Leuven, Departement Sociologie, K.U.Leuven, 2003, 222 blz. + bijlagen.
  
73. VAN DE PUTTE Bart, Het belang van de toegeschreven positie in een moderniserende wereld. Partnerkeuze in 19de-eeuwse Vlaamse steden (Leuven, Aalst en Gent). Leuven, Departement Sociologie, K.U.Leuven, 2003, 425 blz. + bijlagen.
  
74. HUSTINX Lesley, Reflexive modernity and styles of volunteering: The case of the Flemish Red Cross volunteers. Leuven, Departement Sociologie, K.U.Leuven, 2003, 363 blz. + bijlagen.
  
75. BEKE Wouter, De Christelijke Volkspartij tussen 1945 en 1968. Breuklijnen en pacificatiemechanismen in een catch-allpartij. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 423 blz. + bijlagen.
  
76. WAYENBERG Ellen, Vernieuwingen in de Vlaamse centrale - lokale verhoudingen: op weg naar partnerschap? Een kwalitatieve studie van de totstandkoming en uitvoering van het sociale impulsbeleid. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 449 blz. + bijlagen.
  
77. MAESSCHALCK Jeroen, Towards a Public Administration Theory on Public Servants' Ethics. A Comparative Study. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 374 blz. + bijlagen.
  
78. VAN HOYWEGHEN Ine, Making Risks. Travels in Life Insurance and Genetics. Leuven, Departement Sociologie, K.U.Leuven, 2004, 248 blz. + bijlagen.
  
79. VAN DE WALLE Steven, Perceptions of Administrative Performance: The Key to Trust in Government? Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 261 blz. + bijlagen.
  
80. WAUTERS Bram, Verkiezingen in organisaties. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2004, 707 blz. + bijlagen.
  
81. VANDERLEYDEN Lieve, Het Belgische/Vlaamse ouderenbeleid in de periode 1970-1999 gewikt en gewogen. Leuven, Departement Sociologie, K.U.Leuven, 2004, 386 blz. + bijlagen.

82. HERMANS Koen, De actieve welvaartsstaat in werking. Een sociologische studie naar de implementatie van het activeringsbeleid op de werkvloer van de Vlaamse OCMW's. Leuven, Departement Sociologie, K.U.Leuven, 2005, 300 blz. + bijlagen.
  
83. BEVIGLIA ZAMPETTI Americo, The Notion of 'Fairness' in International Trade Relations: the US Perspective. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2005, 253 blz. + bijlagen.
  
84. ENGELEN Leen, De verbeelding van de Eerste Wereldoorlog in de Belgische speelfilm (1913-1939). Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2005, 290 blz. + bijlagen.
  
85. VANDER WEYDEN Patrick, Effecten van kiessystemen op partijssystemen in nieuwe democratieën. Leuven, Departement Sociologie, K.U.Leuven/K.U.Brussel, 2005, 320 blz. + bijlagen.
  
86. VAN HECKE Steven, Christen-democraten en conservatieven in de Europese Volkspartij. Ideologische verschillen, nationale tegenstellingen en transnationale conflicten. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2005, 306 blz. + bijlagen.
  
87. VAN DEN VONDER Kurt, "The Front Page" in Hollywood. Een geïntegreerde historisch-poëtische analyse. Leuven, Departement Communicatiewetenschap, K.U.Leuven, 2005, 517 blz. + bijlagen.
  
88. VAN DEN TROOST Ann, Marriage in Motion. A Study on the Social Context and Processes of Marital Satisfaction. Leuven, Departement Sociologie, K.U.Leuven/R.U.Nijmegen, Nederland, 2005, 319 blz. + bijlagen.
  
89. ERTUGAL Ebru, Prospects for regional governance in Turkey on the road to EU membership: Comparison of three regions. Leuven, Departement Politieke Wetenschappen, K.U.Leuven, 2005, 384 blz. + bijlagen.
  
90. BENIJTS Tim, De keuze van beleidsinstrumenten. Een vergelijkend onderzoek naar duurzaam sparen en beleggen in België en Nederland. Leuven, Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2005, 501 blz. + bijlagen
  
91. MOLLICA Marcello, The Management of Death and the Dynamics of an Ethnic Conflict: The Case of the 1980-81 Irish National Liberation Army (INLA) Hunger Strikes in Northern Ireland. Leuven, Onderzoekseenheid: Instituut voor Internationaal en Europees Beleid [IIEB], K.U.Leuven, 2005, 168 blz. + bijlagen
  
92. HEERWEGH Dirk, Web surveys. Explaining and reducing unit nonresponse, item nonresponse and partial nonresponse. Leuven, Onderzoekseenheid: Centrum voor Sociologie [CeSO], K.U.Leuven, 2005, 350 blz. + bijlagen
  
93. GELDERS David (Dave), Communicatie over nog niet aanvaard beleid: een uitdaging voor de overheid? Leuven, Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2005, (Boekdeel 1 en 2) 502 blz. + bijlagen
  
94. PUT Vital, Normen in performance audits van rekenkamers. Een casestudie bij de Algemene Rekenkamer en het National Audit Office. Leuven, Onderzoekseenheid: Instituut voor de Overheid [IO], K.U.Leuven, 2005, 209 blz. + bijlagen

95. MINNEBO Jurgen, Trauma recovery in victims of crime: the role of television use. Leuven, Onderzoekseenheid: School voor Massacommunicatieresearch [SMC], K.U.Leuven, 2006, 185 blz. + bijlagen