



Faculteit Sociale Wetenschappen

Politieke Wetenschappen

**At a crossroads –
How to change towards
more meaningful performance management?**

Proefschrift voorgelegd tot het behalen van de graad van

Doctor in de Sociale Wetenschappen: Politieke Wetenschappen

aan de Universiteit Antwerpen

te verdedigen door

CORNELIA HOFFMANN

Promotor: Prof. Dr. Wouter Van Dooren

Antwerpen, 2016

Members of the doctoral commission

Prof. dr. Wouter Van Dooren (supervisor), Universiteit Antwerpen

Prof. dr. Koen Verhoest, Universiteit Antwerpen

Prof. dr. Ive Marx, Universiteit Antwerpen

Prof. dr. Geert Bouckaert, Katholieke Universiteit Leuven

Additional members of the doctoral jury

Prof. dr. Peter Bursens, Universiteit Antwerpen

Prof. dr. Jarmo Vakkuri, University of Tampere, School of Management

Cover: Balázs Kajor

This research was financially supported by the Steunpunt Bestuurlijke Organisatie – Slagkrachtige Overheid 2012-2015 (Policy Research Centre - Governmental organization - Decisive Governance 2012-2015).

Copyright © Cornelia Hoffmann, 2016

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means without prior permission of the author.

ISBN: 9789057285097

Depot number: D/2016/12.293/5

Printed by Universitas, Antwerp, Belgium

ACKNOWLEDGEMENTS

*It is good to have an end to journey toward,
but it is the journey that matters, in the end.
~ Ursula K. Leguin*

My personal journey from accepting the PhD position at the University of Antwerp in March 2012 until the defence of my thesis in March 2016 has been a memorable adventure. The tangible result sits in your hands right now (or on your screen, for that matter). However, this result would not have been possible without the many intangible hands that helped me on this journey in so many ways: to encourage me to walk a path without knowing exactly what the end of it would bring, to dare to be lost sometimes, and, most importantly, to keep going.

My thanks go first of all to my supervisor Wouter Van Dooren, who supported me not only with his knowledge and expertise, but also with his motivation and enthusiasm – “het moet plezant blijven” (“it has to be fun”) is certainly a good motto to keep in mind when embarking on such a journey. I am also grateful for the support and feedback from my doctoral commission, Koen Verhoest, Ive Marx, and Geert Bouckaert, as well as from the external jury members Peter Bursens and Jarmo Vakkuri. Jarmo, I especially appreciate your having given me the opportunity to work at the School of Management of the University of Tampere for three months. It was a rewarding experience to work with you and the other team members. I am particularly obliged to Anniina Autero for her help in finding and contacting the Finnish interviewees and her always kind and encouraging words.

As a big part of this thesis is based on interviews in Finland and Lithuania, I want to express my gratitude to all the respondents who took the time to tell me about their experiences in their respective organisations. Without their willingness to share their thoughts with me, my thesis would not have been possible in this form. Special thanks in this regard also to Inga Kirstukaite, who opened the doors to the Lithuanian administration for me.

No (PhD) journey without money: I certainly owe a debt of gratitude to the Flemish Government for funding my research within the scope of the Steunpunt Bestuurlijke Organisatie – Slagkrachtige Overheid 2012–2015 (Policy Research Centre – Governmental organization – Decisive Governance).

I need to add a few words about my experience as consultant at the OECD in Paris (September 2015–January 2016): What might at first have looked – though interesting and inspiring – like a detour, later proved to be a decisive step toward the completion of my thesis, as it was during this period that I came to know the ins and outs of the

Lithuanian administration. Thanks to Eva Beuselinck for having been such an encouraging, forward-looking, and structured ‘mentor’.

Neither did I undertake this journey by myself. It is also thanks to many colleagues (who in the meantime also became friends) that I have reached my destination. My gratitude goes to my ‘coffee buddies’ Jan and Martijn, to Chiara, to Astrid and Margaux, and all the ‘PAM young stars’, as well as to the PANDAs (Astrid, Chiara, Jan, Seb, and Sjors) for all their encouragement, support, and constructive reflections along the way. Thanks to Peter: we accomplished our pact. All of you made the trip worthwhile.

Last but by no means least, I want to mention my family and other friends (not work-related) who accompanied me on this journey, each in their own way, and especially Balázs, who met and bore with me in probably the most stressful period. Köszönöm szépen. Thanks for everything.

Antwerp, March 2016

Table of Contents

1. Introduction	9
1.1. Setting the scene	9
1.2. The core underlying concepts: Performance and performance information, performance measurement, performance management	12
1.3. Structure and approach.....	14
2. Dysfunctional effects of indicators.....	17
2.1. Different potential effects of performance information use from a general perspective	18
2.2. Empirical studies of the effects of performance information in the public sector.....	23
2.3. Organisational and political contexts.....	28
2.4. Tension between internal improvement and external reporting.....	30
2.5. Conclusion and implications.....	34
3. Use of performance information	35
3.1. The different uses and purposes of performance information.....	35
3.2. Accountability in performance management systems	39
3.2.1 Accountability – an approximation to the concept	39
3.2.2 The accountability paradox.....	42
3.3. Learning in performance management systems.....	44
3.3.1 Organisational learning – an approximation to the concept.....	45
3.3.2 Types of organisational learning: Single-loop vs. double-loop learning ..	47
3.3.3 The learning organisation – an approximation to the concept.....	50
3.3.4 Context matters: Structure and culture	54
3.4. Accountability <i>and</i> learning in performance management systems?	57
3.5. Accountability, learning, and trust	61
3.6. Implications of the use of performance information for the design of performance measurement systems	64

3.7. Conclusion and implications.....	67
4. Methodological approach.....	71
4.1. A “most likely/least likely” comparative case study – the what and the why...	71
4.2. Analytical approach: Qualitative interviews – the how.....	73
4.3. Reflections on the approach chosen – from reliability, validity, and generalisability to trustworthiness and exemplary knowledge	78
4.3.1 From reliability and validity to trustworthiness.....	79
4.3.2 From general to exemplary knowledge (<i>phronesis</i>).....	80
5. Mapping out the terrain: Finland and Lithuania in context	83
5.1. A politico-administrative description of Finland.....	83
5.2. The Finnish performance management system – “an agreement-based control system”	84
5.3. Current situation and challenges	86
5.4. A politico-administrative description of Lithuania.....	89
5.5. The Lithuanian performance management system – “a formalised system for performance monitoring and reporting of results”	90
5.6. Current situation and challenges	92
6. In the field: The Finnish and the Lithuanian systems analysed	95
6.1. Use of performance information in Finland.....	95
6.1.1 On the purpose of the performance management system	95
6.1.2 On the design of the performance management system.....	97
6.1.3 On the use of performance information	99
6.1.4 On the reporting of performance information	101
6.1.5 On the structure and culture of the performance management system	103
6.1.6 On learning forums	105
6.2. Finland – torn between accountability and learning?	108
6.3. Use of performance information in Lithuania	110
6.3.1 On the purpose of the performance management system	110
6.3.2 On the design of the performance management system.....	111

6.3.3	On the use of performance information	113
6.3.4	On the reporting of performance information	116
6.3.5	On the structure and culture of the performance management system	119
6.3.6	On learning forums	122
6.4.	Lithuania – a clear case for accountability?	124
6.5.	Conclusion and implications	127
7.	Conclusion and outlook	131
7.1.	Findings and theoretical contribution	131
7.2.	Limitations of this study and future research ideas	135
7.3.	Practical implications	137
Question 1	Where are we going with performance management systems? ...	137
Question 2	Who gains and who loses; by which mechanisms of power?	138
Question 3	Is this development desirable?	139
Question 4	What should be done?	140
8.	Bibliography	143
	Samenvatting (Dutch summary)	159

List of tables

Table 1	Three uses of performance information	38
Table 2	Performance accountability and organisational learning compared	38
Table 3	Forums and processes of accountability	40
Table 4	Traditional and innovative accountability compared	60
Table 5	The design of the performance measurement system according to its use.....	67
Table 6	Overview of organisations and respondents interviewed.....	74
Table 7	Main differences in focus between the old governance approach and the future approach.....	88
Table 8	Comparison of findings for Finland and Lithuania	128
Table 9	Propositions and findings	129

List of figures

Figure 1	Performance target paradox - the tendency towards a target.....	32
Figure 2	Types of knowledge use and purpose	36
Figure 3	Eight purposes of use of performance information	36
Figure 4	Organisational learning and use of performance information	44
Figure 5	The structure-behaviour relationship	52
Figure 6	Distinctions between the four aspects/types of the learning organisation.....	53
Figure 7	The structure-culture-behaviour relationship.....	56
Figure 8	Performance prism	85
Figure 9	Strategic planning system in Lithuania	91
Figure 10	Trade-off between accountability and learning	134

List of boxes

Box 1	Effects of the use of performance information	27
Box 2	Elements of learning forums	55

1. INTRODUCTION

*“Sal, we gotta go and never stop going 'till we get there.’
‘Where we going, man?’
‘I don't know but we gotta go.’”
~ Jack Kerouac, On the Road*

1.1. Setting the scene

From “made in Germany” to “faked in Germany”, the scandal over Volkswagen (VW) manipulating software in diesel engines to improve performance results when tested for pollution emission levels has provoked international outcry. VW, known for its diesel technology and comparable low prices, apparently could no longer compete in the market and reacted by breaching environmental standards (Schulz, 2015). This manipulation had an impact beyond VW itself; manipulating the measurement of CO₂ emission levels also affected the assumed fuel consumption. Thus, car owners thought they would use less fuel than they actually did. In turn, this had an effect on the motor vehicle tax that is based on the fuel consumption level, meaning that the state might have received fewer tax revenues. In addition, CO₂ emissions are a major contributor to the greenhouse effect and, thus, it presumably affected public life. As allowed CO₂ emission levels are furthermore regulated by the EU, this manipulation also infringed EU regulations. In summary, the impact of the manipulation of “performance results” in the private sector spilled over into the public domain (Die Zeit, 2015).

This cheating scandal reveals the complex environment within which organisations currently find themselves, notably the pressure to perform, and the consequences manipulative behaviour can have in response to this pressure (Moynihan et al., 2011; Vakkuri & Meklin, 2006). Organisations have to satisfy a variety of actors with the provision of services, and are held accountable by the same actors. However, often the actors involved have different or even contrasting interests or goals. How is it possible to satisfy all of them simultaneously?

The economic and political scientist, Ayad Al-Ani, explains that VW’s behaviour is symptomatic of organisations in which members are led by fear of not being able to reach their goals, fear of disappointing others, and consequently, of being kicked out of the game (Al-Ani, 2015). This, however, is based on a lack of “new procedures, innovative technologies, and collaborative organisation forms to solve problems also in a ‘new’ way”. Organisations are “left with known strategies, which are more often than not insufficient and must be maxed out to the bitter end and sometimes beyond” (Al-Ani, 2015).

The aforementioned example of VW seems to be the result of the high pressure to perform that the company experienced. In fact, Morieux (2011) claims that companies

now have six times more performance requirements than they did 60 years ago. Today, CEOs have to comply with 25 to 40 performance targets on average, as opposed to the four to seven targets at the beginning of performance measurement in the private sector.

The situation does not look much different in the public sector. Since the introduction of New Public Management (NPM) in the 1980s – the doctrine that aimed at making the public sector more efficient and effective by copying management practices from the private to the public sector – governments and public administrations have introduced manifold performance targets, indicators, balanced scorecards, and (international) rankings (Van Dooren, Bouckaert & Halligan, 2015). Indeed, “economy, effectiveness, efficiency” and “doing more with less” have become central mantras for characterising the NPM (Kettl, 2002; Osborne & Gaebler, 1992).

However, an increasing number of critical voices question the basic assumptions of performance management (Radin, 2006). The “holy trinity ... of efficiency [i.e. clarity, measurement, accountability] make[s] human efforts derail”, claims Morieux (2015) in a fairly recent TED Talk.

W. Edwards Deming, known for his influence on the quality management movement (and his subsequent turn against it), goes even so far to claim the following:

“Our prevailing system of management has destroyed our people. People are born with intrinsic motivation, self-respect, dignity, curiosity to learn, joy in learning. The forces of destruction begin with toddlers – a prize for the best Halloween costume, grades in school, gold stars – and on up through the university. On the job, people, teams, and divisions are ranked, reward for the top, punishment for the bottom. Management by Objectives, quotas, incentive pay, business plans ... cause further loss, unknown and unknowable.” (quoted in Senge, 2006, p. xii).

The list of critiques continues. The intentions behind NPM-inspired management techniques were noble; hopes for success were great. Where did their noble intentions go awry?

In fact, results-based management, performance-related pay, and other forms of tight performance management systems have provided a route to gaming and cheating, in the form of data or output manipulation. Thus, it is widely recognised in the academic literature that administrative, indicator-based accountability might have missed its target (Arndt & Oman, 2006; Bevan & Hood, 2006; Hood, Dixon & Beeston, 2008; Luts, Van Dooren & Bouckaert, 2008; Pollitt, 2013; Van Dooren & Van de Walle, 2011; Van Dooren et al., 2015). Various scholarly contributions show the problems of finding appropriate indicators (Arndt & Oman, 2006) and of international rankings based on performance indicators (Luts, van Dooren & Bouckaert, 2008; Hood, Dixon & Beeston, 2008). The negative side effects of indicators and performance management systems, such as gaming and cheating, have received ample scholarly attention (Bevan & Hood, 2006; Pollitt, 2013; Van Dooren & Van de Walle, 2011; Van Dooren et al., 2015; Van Thiel & Leeuw, 2002). At the same time, accountability is vital for democratic systems and, as such, unlikely to disappear.

Accountability takes place in different forums: the political, the administrative, the judicial, the public, and the market (Romzek, 2000; Van Dooren & Willems, forthcoming). Indicator-based accountability is generally designed for the administrative forum. Actors are held to account by means of indicators in performance contracts, performance budgets, or scorecards. Nevertheless, these indicators seldom remain in the administrative forum only; actors from other forums (e.g. politicians, journalists, consultants etc.) can pick up these indicators at any time and use them for their own purposes. Consequently, instead of leading to organisational improvement and innovation, indicator-based accountability has led to risk avoidance and seeing indicators as an end in themselves, as opposed to a means to an end (Van Dooren & Willems, forthcoming).

Hence, increasing concerns have been voiced about accountability-based performance management systems, especially when accountability is based on indicators (Moynihan, 2005; Olejniczak & Newcomer, 2014; Van Dooren & Willems, forthcoming). As a potential way out, recently scholars increasingly suggest to direct attention to learning purposes (Courty & Marschke, 2007; Greiling & Halachmi, 2013; Halachmi, 2002; Kettl, 2001; Mayne, 2007; Molleman & Timmerman, 2003; Olejniczak & Newcomer, 2014; Schillemans, Twist & Vanhommerig, 2013; Van Dooren & Willems, forthcoming; Yang & Holzer, 2006). Instead of holding actors to account for their results, they are to be held accountable for the processes in which they are involved. As such, the focus shifts towards quality of actions, as opposed to failure (Lewis & Triantafillou, 2012). Furthermore, rather than engaging in hierarchical, often external, accountability relationships, actors would engage in horizontal, internal, learning efforts, and share the responsibility for their actions.

Nevertheless, these approaches have received some criticism. It is argued that learning purposes are likely to supplement, rather than replace accountability for performance. Hence, it would also require the collection of more data, rather than less and, consequently, be more costly for the organisation at stake. It further requires significant participation and dialogue, and might lead to change in the name of improvement, as opposed to when it is truly necessary (Lewis & Triantafillou, 2012).

Are performance management systems at a crossroads, with accountability and learning purposes leading in two different directions? Alternatively, is there a middle-road that could satisfy both ends? Empirical investigations on this subject are scarce. The (organisational) learning literature has so far focused mainly on the private sector, and thus, does not offer many insights into accountability challenges within the public sector. The accountability literature, on the other hand, pays more attention to performance management procedures and less to organisational learning. Thus, there exists little understanding of whether and how learning and performance accountability can coexist (Olejniczak & Newcomer, 2014).

To consider these questions, we must closely examine the central element in the debate – the use of performance information, and it is on this that the thesis focuses. The underlying premise of this thesis is that *the use of performance information for*

accountability purposes is unlikely to be compatible with using it for learning purposes. This is in particular due to the different accountability forums in which performance information is, or can be, used.

Against this background, the following research questions are addressed:

How do performance management systems deal with the tension between accountability and learning purposes of the use of performance information?

- 1) What are the implications for performance management systems of the extent to which accountability elements are used?
- 2) What are the implications for performance management systems of the extent to which learning elements are used?
- 3) Are there indications of incompatibilities between the two?

Now that the context and the purpose of this dissertation are presented, I examine more closely the main concepts applied throughout this study. This is carried out in the following section.

1.2. The core underlying concepts: Performance and performance information, performance measurement, performance management

This section briefly defines the core underlying concepts of this dissertation. These are performance, performance measurement and performance information, and performance management. Accountability and learning, the central concepts for the empirical part of this thesis, deserve special attention and are therefore analysed and discussed in detail in chapter 3.

Performance as a concept or idea has become ubiquitous to the extent that it has already developed into a “movement”, argues Radin (2006). It is thus all the more important to explore what this movement entails before following it blindly. To do so, first the main concepts inherent in the “performance movement” are briefly explained.

Performance: Performance is both a concept and an agenda. As concept, it describes “sustainable results”, which manifest themselves in the “quality of actions and the quality of achievements”. Thus, an organisation that performs well is an organisation that has “the capacity to perform and converts this capacity into results – outputs and outcome” (Van Dooren et al., 2015, p. 4). As an agenda, performance entails “a programme of change and improvement” (Van Dooren et al., 2015, p. 4).

Performance information: Performance information is the result of “deliberate activities for quantifying performance” (Van Dooren et al., 2015, p. 32). The most obvious form of performance information is performance indicators. They are generally categorised according to the sort of information they entail and include inputs, outputs,

or results, or outcomes or impact. Inputs usually include financial and human resources and are allocated to reach certain outputs. The input-output ratio is what – in economic terms – we call “efficiency”. Outputs are the “immediate results of government activity”, whereas outcomes describe the “final impacts of these activities” (OECD, 2009, p. 10). The output-outcome ratio is measured by the term “effectiveness”.

However, while governments can be held to account for the outputs of their policies, this is less evident in the case of outcomes, as the latter are also influenced by external factors, which are not under the control of governments (OECD, 2009). Furthermore, in the private sector, the outcome (seen as value added) can be assessed according to “private value”. In the public sector (where we talk about “public value”) this is much more difficult, since outcomes cannot usually be assessed in monetary terms (Rainey, 2009; Van Dooren, De Caluwe & Lonti, 2012, p. 493).

Consequently, we need a more substantial approach to performance, one that includes public values, rather than the purely analytical concept presented above.

Performance as public values: Based on the framework of Hood (1991, pp. 11–14), Van Dooren et al. (2012, pp. 494–495) distinguish three types of performance. Firstly, product performance focuses on government being lean and purposeful, i.e. “matching resources to defined tasks”. Economy, efficiency, and effectiveness of government are the key terms. Secondly, process performance emphasises fairness and honesty, which the government has to provide. Thirdly, regime performance stresses robustness and resilience of the public sector and, as such, primarily entails an *a posteriori* element. Only after something (a “failure”, “worst case scenario”) has occurred, can it be measured whether the system has been sufficiently robust and resilient.

The NPM-movement has particularly been criticised for having focused too much on product performance values (efficiency, effectiveness) at the expense of other values such as equity and professionalism (Pollitt & Bouckaert, 2011; Radin, 2006; Van Dooren et al., 2015). Robustness and resilience of the public sector have been discussed in the light of the financial and fiscal crisis or, recently, in relation to the ability to adequately respond to and cope with the refugee crisis.

Performance measurement: Performance measurement is “the bundle of activities aimed at obtaining information on performance”. This includes measurement based on performance indicators, but also information received via external advice and consultants and citizens (Van Dooren et al., 2015, p. 20).

Performance management: Performance management is “a type of management that incorporates and uses performance information for decision-making” (Van Dooren et al., 2015, p. 20).

Performance management systems are thus systems aimed to support decision-making based on performance information, which is measured, analysed, and evaluated. As such, performance measurement is not the same as performance management – but performance management always includes some form of performance measurement (Kettl, 2001).

1.3. Structure and approach

I proceed in **chapter 2** with a review of the literature on the effects of performance information with the aim to find the underlying mechanisms for the occurrence of negative or unintended effects. This leads to the need for a closer examination of the different ways in which performance information can be used. After a general, broad overview of the possibilities and opportunities for the use of performance information in **chapter 3**, I focus on two forms of use, which can be described as the two ends on a continuum of uses: Accountability and learning. While some claim that these are opposites, others contend that there are possibilities for synergies and that one might even feed the other. Furthermore, the use of performance information also plays an important role in the design of performance measurement systems. This is also addressed in this chapter.

To analyse this issue empirically, I compare two cases in a most likely / least likely research design. **Chapter 4** explains the methodological and analytical choices for the empirical part in detail. The most likely case is the Finnish performance management system, where, based on its administrative structures and culture, I would expect fertile ground for shifting away from an accountability and control system towards a system based on learning and dialogue. In contrast, the Lithuanian performance management system serves as a least likely case. By conducting a comparative case study, I want to determine to what extent performance information is used for accountability and/or learning purposes, and what the implications are for each outcome.

A qualitative approach such as this has several advantages, which is also highlighted in the methodological section of this thesis (**chapter 4**). For example, it allows me to thoroughly analyse my cases by bringing together different points of view. I am thus able to acquire first-hand information by talking to the people involved directly, rather than by means of surveys or document and literature analyses. Furthermore, I am able to pay due respect to the context in which the study is situated.

It is this context-sensitivity that, on a broader level, is an essential part of what Flyvbjerg (2001) calls “*phronetic* social science”. The term *phronesis* was introduced by Aristotle, who distinguished between scientific knowledge (*episteme*), technical knowledge or expertise (*techne*), and practical wisdom (*phronesis*). Flyvberg (2001) argues that social sciences have focused too much on (and at the same time have been reduced to) *episteme* and *techne*, at the neglect of *phronesis*. However, it is exactly here, where social sciences are the strongest. As such, *phronetic* social science recognises “the power of example” by placing “context, judgement, and practical knowledge” centre stage (Flyvbjerg, 2001, p. 24). Consequently, Flyvbjerg (2001) and Flyvbjerg, Landman & Schram (2012) consider case studies as an excellent – but by no means the only – way to conduct *phronetic* research. *Phronetic* social science attempts to provide answers to the following four value-rational questions:

- 1) Where are we going?
- 2) Who gains and who loses; by which mechanisms of power?

- 3) Is this desirable?
- 4) What should be done?

This dissertation is written with these questions in mind. Answers to these questions are provided in the concluding chapter of this thesis.

Chapter 5 outlines the context in which the two case studies are conducted, and depicts the case-selection strategy in greater detail. In particular, the structure of the performance management systems of the respective administrations, their developments, and current challenges are presented. In this context, also the administrative culture plays an important role.

Against this background, **chapter 6** constitutes the actual empirical analysis of the cases. The findings are presented and analysed against the underlying premise of the thesis. The last section of this chapter compares the two cases and suggests the implications of the findings.

The **concluding chapter** summarises the main findings and elaborates on the implications for performance management systems in general. Bearing in mind the rationale of *phronetic* social science, this part of the dissertation attempts to develop “partial answers” to the four value-rational questions posed above, hoping to provide “input to the ongoing social dialogue about the problems and risks we face, and how things may be done differently” (Flyvbjerg, 2001, p. 61).

2. DYSFUNCTIONAL EFFECTS OF INDICATORS

“If you know the unintended consequences of your acts, you are in a better position to analyse the facts.”

~ Tasneem Hameed

In an article on the performance report of the Dutch police, De Koning (2015) questions the measurability of the performance of the police. He claims that the bureaucratic “to measure is to know” mentality (“meten is weten”) leads to a focus on reaching targets and goals, while the focus on the bigger picture, quality, craftsmanship, and professional ethics gets lost.

In Atlanta, the “largest cheating scandal in the nation’s history” was brought before the court: In 2015, 11 teachers were found guilty (and sentenced to prison) for partaking in standardised test cheating, mostly by manipulating test scores with the aim to raise the test scores, and as such increase the schools’ and students’ achievements, as well as the teachers’ bonuses (Blinder, 2015).

In the US, the veterans’ health care centres have been blamed for “juggling appointments and sometimes manipulating wait lists to disguise long delays for primary and follow-up appointments”, which ultimately resulted in very poor patient care and even deaths, allegedly. “The performance data the VA puts out is garbage — it's designed to make the VA look good on paper. It's their 'everything is awesome' approach,” the chief of psychiatry at the VA St. Louis Health Care System is quoted. “There's a 'don't ask, don't tell' policy. Those who ask tough questions are punished, and the others know not to tell.” (Zucchini, Carcamo & Zarembo, 2014).

These are just a few, recent examples of “performance perversity” (Moynihan, 2014), which call for reflection on performance management and the use and usefulness of performance information. “The ultimate test of an effective performance management system is whether it is used, not the number of goals and measures produced”, argue Moynihan & Lavertu (2012, p. 594). However, at the same time, Dahler-Larsen (2014) claims that evidence for “explicit intentional effects” of such performance information use is scarce. Evidence for unintended or negative effects, on the other hand, is plentiful.

The present chapter aims to find an answer to the questions of what these effects are and why (and under which mechanisms) the use of performance information results in dysfunctional effects. It does so by providing an overview of the literature addressing the effects of the use of performance information. For this aim, articles of various public administration journals (*Journal of Public Administration Research and Theory*, *Public Administration Review*, *Public Management Review*, *Public Money & Management*, *Public Performance and Management Review*, and its predecessor, *Public Productivity and Management Review*, *International Journal of Public Administration*, *International Journal of Public Sector Management*, and the *International Journal of Productivity & Performance Management*) have been analysed. The articles were selected by applying a combination of database search and the snowball principle. For the database search, key words and their combinations, as well as the concrete journal titles, were inserted in the

Web of Science database search engine. These key words included the following: performance measurement, use, effect, indicator, and impact. The time span for the search results was limited to 1995-2015. The year 1995 seemed to be an appropriate starting point for the research because, after the introduction of performance management systems in the scope of the NPM movement in the 1980s and 1990s, sufficient time had passed to be able to articulate the first results and experiences with the different forms of performance management systems. Nevertheless, through the additional snowball principle, whereby references in relevant articles were tracked backwards, a few studies deemed relevant and important, which were dated prior to 1995, were found and included in the analysis. Another selection criterion was the language. Only English-speaking articles (and journals) were included. While being a pragmatic choice, this choice also reflects the fact that most European scholars publish in English journals. In addition, only journal articles are included in the literature review.

The literature on the effects of the use of performance information can be divided into different groups. First, there are articles that describe the different potential effects of performance information use from a general perspective. The contributions of Behn and Kant (1999), Berman (2002), Bouckaert and Balk (1991), Grizzle (2002), de Bruijn (2002), Hood (1974), Hood (2012), Pidd (2005), Smith (1995), and Van Thiel and Leeuw (2002) belong to this group. Furthermore, there are those who analyse empirically concrete cases or sub-sectors of the public sector. Most prominently analysed are the (British) health care system and the education system (Bevan & Hood, 2006; Bohte & Meier, 2000; Courty & Marschke, 2007; Heinrich, 1999; Jung & Lee, 2013; Kelman & Friedman, 2009; Nielsen, 2014; Nõmm & Randma-Liiv, 2012). While most articles describe the effects from an outside perspective, a few do include the insiders' points of view, such as the perception of public employees (Jung & Lee, 2013). A further important, relatively large group of authors research the organisational and political contexts to arrive at more nuanced articulations about the effects of performance information use (Berman & Wang, 2000; Dahler-Larsen, 2014; Hood, 2012; Kerpershoek, Groenleer & de Bruijn, 2014; Noordegraaf, 2008; Noordegraaf & Abma, 2003; Taylor, 2009). An interesting perspective in this context is provided by Gregory and Lonti (2008) who address the symbolic versus the practical significance of unintended, or negative, effects. Finally, the tension between internal improvement and external reporting is addressed by several authors (Alcock, 2004; Bouckaert & Peters, 2002; Goddard & Mannion, 2004; Taylor, 2009).

2.1. Different potential effects of performance information use from a general perspective

Let us commence with the articles on the effects of performance information use from a more general perspective. Berman (2002) questions to what extent performance management is useful. He claims, "empirical evidence suggests that such [successful] uses of performance measurement are the exception rather than the rule." More

investment in information technology (IT) systems and human resources would be necessary to collect data that is more useful, by which he refers to data that is timelier and reflects on outputs. Nevertheless, he remains somewhat optimistic in that he prophesies that performance measurement will need another “7 to 10 years” until we will be able to draw on its full potential. Now, 13 years after this estimation, the question of the usefulness of performance measurement is just as relevant as it was then. However, Berman's (2002) estimation might have been a little too optimistic, as the number of recent criticisms of performance measurement indicates that is examined below.

The challenges of performance measurement and management are often described metaphorically in terms of “diseases”. As early as in 1974, Hood referred to administrative diseases and dysfunctionalities in public administrations in general (Hood, 1974). Almost 20 years later, Bouckaert and Balk (1991) revisit the notion of diseases, and discuss diseases and cures – this time within the context of public productivity measurement. They divide the “diseases”, that is, dysfunctional effects, of performance measurement into three groups:

The first group concerns the assumptions of measurement. The diseases of this group describe the problem of attempting to measure something despite the fact that there seems to be no need for it (Pangloss disease), the complexity of the public sector, which makes it impossible to adequately measure it (impossibility disease), and the problem of not admitting that things go well (hypochondria).

The second group describes diseases of perceived numbers and volumes. This includes over- or underestimation of measured volumes (convex/concave disease), “measurement itself stimulat[ing] the unwanted production of output” (hypertrophy), “measurement caus[ing] a decrease in the [output] being measured” (atrophy), and the issue with details in measurement; the more detailed the measurement is, the higher the results (Mandelbrot disease).

The third group captures the content, position, and number of measures. Mixing input, output, and outcome indicators (measurement pollution), “needless mushrooming of measurements” (inflation), imposition of measures by external stakeholders or top-level management without consulting lower level(s) (enlightenment/top-bottom), focus on the short run at the expense of medium- or long term focus (time-shortening), indicators and measures not fitting together (mirage), and measures and goals or purpose contradicting each other (shifting), belong to this group. Bouckaert and Balk (1991) conclude that close attention must be paid to the impact of (the use of) performance measurement in an organisation.

Stressing the importance of the design and the implementation of the performance management system, Bouckaert and Peters (2002) address the dilemmas and consequences of performance-based management. Potential pitfalls include the fact that costs of performance measurement (and management) are more visible than (potential) benefits. Moreover, there is a risk of focusing excessively on the more easily measurable lower levels of performance (myopic, “frog view”), as opposed to the more complex and

more important societal matters, which, however, are more difficult to measure. In addition, quality is often measured in relation to quantity, i.e. in relation to costs. What is to be done in the case where a service can be improved by making it accessible to more people, but only by lowering its quality? This leads to the next dilemma – performance and resource allocation. Should higher performance be rewarded with more resources? However, would that in fact lead to more efficiency? Alternatively, should resources be cut to increase efficiency even further? What is to be done in cases of under-performance? Should resources be cut as a form of sanction, while increasing the likelihood for even poorer performance for the subsequent reporting period? Perhaps increased funding will increase the likelihood for success in the next period. However, would that not provide an incentive for “structural” under-performance to increase funding? Similarly, best practices can have both a motivational and a de-motivational effect. The latter could be the case for organisations for which it is simply not possible (due to their structure or tasks) to create best practices. They would thus never be rewarded. Hence, the authors conclude that the design and implementation of performance management systems has to be considered carefully and investigated thoroughly.

Another general account of negative, unintended consequences of performance measurement is provided by Grizzle (2002). She provides examples of unintended consequences when measuring outcomes or results, customer satisfaction, quantity of work done, and efficiency. These include “teaching the test” in public schools (see also Wiggins & Tymms (2002) on the effects of using league tables in English schools versus performance of Scottish schools without league tables), downgrading offenses in crime reports, creaming (that is, including the individuals or target groups in the measurement who are most likely to increase the performance score), answering customer satisfaction surveys under (indirect) pressure of superiors, definition creep, and lowering quality. To prevent or ameliorate these unintended consequences, Grizzle (2002) emphasises the indispensable role trust plays in the game. Employees need to trust that performance information will be used in a fair manner. Furthermore, designing the measurement process in such a way that it is considered an empowering experience for employees rather than discouraging, could prevent deliberate negative consequences. This could start with creating favourable working environments for employees. Finally, she encourages reconsidering the basics on which performance measurement systems are generally built; perhaps it is not necessary, or feasible, to measure all programmes in terms of efficiency, amount produced, and cost-effectiveness.

Hood (2012) examines “public management by numbers as a performance-enhancing drug”. He conducted a literature review on the question of what makes management by numbers a performance-enhancing drug in some contexts, but not in others. He finds that the performance-enhancing use depends on the purpose for which numbers are used. Thus, Hood distinguishes between using performance information for targets, rankings, and “intelligence”. With “intelligence”, he refers to using performance information as “background information for choice by users or for policy change or for management intervention” (Hood, 2012, p. S86). However, all these performance-

enhancing effects accompany performance-obstructing effects. Consequently, trade-offs have to be considered when opting for one purpose or the other. In addition, these three types of management by numbers do not work equally well in all contexts. Hence, careful attention has to be paid to the organisational culture (at operating, policy-making, or system-design levels) if management by numbers is to work in a performance-enhancing way.

While Bouckaert and Balk (1991) analyse dysfunctional effects from a more general perspective, Behn and Kant (1999) address them in a concrete example, i.e. performance contracting. They find three groups of pitfalls: 1) Punishment, 2) revolving door, and 3) complexity. Regarding punishment, the authors contend that performance contracting may inhibit experimentation, encourage innovation in cost cutting, but not in service delivery, and restrain overachievement. As concerns the revolving door effect, they argue that performance contracting would not provide for start-up costs, would inhibit symbiotic relationships, and would reward promises, not performance. Finally, complexity describes the fact that performance contracting must rely on outputs instead of outcomes, uses measures that can distort behaviour, might encourage creaming or cherry picking, and it may undermine equity and fairness. However, their analysis does not stop here. Eight strategies are proposed to overcome these potential pitfalls of performance contracting. The strategies are the following (Behn & Kant, 1999, pp. 479–483):

- 1) Understand the relationship and difference between the mission and the measures; ...
- 2) Create contracts based on outputs that are (a) linked to the mission; (b) easy to measure, understand, and reproduce; and (c) facilitate benchmarking; ...
- 3) Start simple and ratchet up; ...
- 4) Monitor frequently lots of indicators of performance, not just the one(s) specified in the contract; ...
- 5) Be prepared to learn, change and improve; ...
- 6) Work collaboratively – not adversarial – with contractors; ...
- 7) Pay vendors not just for the final output but also for significant, well-defined progress; ...
- and 8) favour vendors with a track record.

While getting into detail about the nuts and bolts of performance contracting is beyond the scope of this dissertation, the strategies are equally important and valid for the general design and use of performance management systems. These issues are revisited in the following analysis of articles on the concrete examples of performance measurement effects.

Smith (1995) departs from the fact that performance measurement systems are, in fact, control systems. Improving organisational efficiency, an often-stated goal and purpose of performance measurement systems, is only one form of such control. The author identifies eight unintended consequences, which he illustrates by providing examples from the British public sector. These include a tunnel vision, resulting from emphasis on quantifiable performance measures at the neglect of unquantifiable ones; sub-optimisation, which refers to the “pursuit of narrow local objectives by managers, at the expense of the objectives of the organisation as a whole.”; myopia, the “pursuit of short-

term targets at the expense of legitimate long-term objectives.”; and measure fixation, i.e. the “emphasis on measures of success rather than the underlying objective.” Misrepresentation, referring to the “deliberate manipulation of data so that reported behaviour differs from actual behaviour” – “creative reporting and fraud” are the most notable forms of such behaviour. Misinterpretation then describes the likely effect resulting from the complexity of issues at stake and the bounded rationality of the controllers. Gaming, another well-known effect, is the “deliberate manipulation of behaviour to secure strategic advantage.” Finally, ossification describes the “organisational paralysis brought about by an excessively rigid system of performance evaluation”.

Some strategies to overcome these negative effects include involving staff at all levels of the organisation in the design and implementation of the performance measurement system, keeping performance indicators flexible, and reviewing the system constantly. Interestingly, Smith (1995) also suggests quantifying every objective “however elusive” (p. 304) as a potential strategy to mitigate unintended consequences. This is surprising, as the quantification of measures is also one of the negative effects (tunnel vision). Nevertheless, the former three suggestions are also reflected by other researchers (e.g. Behn & Kant, 1999; Bouckaert & Peters, 2002; Fryer et al., 2009; Van Dooren, 2011) and thus seem indeed to be a viable option.

In fact, it is certain that the story about performance management is not a simple one. Van Thiel and Leeuw (2002) describe it as “performance paradox” which refers to the “weak correlation between performance indicators and the performance itself”. As such, it deals with the reporting on performance, rather than performance itself (Van Thiel & Leeuw, 2002, p. 271). The authors distinguish between an unintended and a deliberate paradox. At the basis of the unintended paradox are unintended effects of the use of performance information. Van Thiel and Leeuw (2002) list the following unintended effects:

- Increased monitoring costs of the organisations and the state
- Dysfunctional effects, such as tunnel vision, sub-optimisation, lack of innovation
- Symbolic behaviour
- Uncertainty about what is being measured.

The reasons for the appearance of these effects are various, such as minimal accountability requirements, elusiveness of policy objectives, non-quantifiable policy goals, and a strong emphasis on monitoring and efficiency of an organisation (Van Thiel & Leeuw, 2002). The latter is especially interesting, as increasing monitoring and efficiency is said to be one of the primary goals of introducing performance management in the first place (Osborne & Gaebler, 1992). Thus, here we have our paradox. Performance management is supposed to increase efficiency of an organisation, but in fact, by measuring it, we might create the opposite effect.

The deliberate performance paradox then builds on the conditions for the unintended paradox. Known practices include cream skimming or cherry picking, and myopia.

2.2. Empirical studies of the effects of performance information in the public sector

Moving from the more general to the more concrete, Heinrich (1999) analyses the extent of effective use of performance-based contracting in a local job-training agency. She concludes that the design of the performance standards system could be improved, as currently performance measures are not strongly aligned with programme goals. Furthermore, cost-per-placement questions play a predominant role at the expense of (improved) service quality. Consequently, Heinrich contends that the focus should be shifted towards assessing the facilitation of achieving the organisation's *central* goals.

In a similar vein, Courty and Marschke (2007) analyse the performance measurement system in a US government job training programme for which a federal agency, the US Department of Labour, sets performance measures for local agencies. The authors conclude that employees involved in this programme changed behaviour when the measurement rules changed. For instance, the Labour Department set a 90-day rule, which called for the agency to measure enrolled people's performance upon termination of their training, while in case a participant had not received training for 90 days this enrolment should be brought to a close. The agencies' reaction, however, was to maximise performance scores by ending enrolment of their participants, not after they finished the training but once they were employed. This, according to Courty and Marschke (2007), increased the employment rate outcome by 20 %. The authors suggest that this type of gaming and cheating behaviour is a result of information asymmetry between those who set the performance targets and those whose performance is being measured. Consequently, Courty and Marschke (2007) call for a continuous trial and error feedback cycle. Performance measurement has to be seen as a dynamic process in which measurement and targets have to be monitored and revised constantly. A method to avoid this sort of gaming behaviour could be to randomise measurement (while maintaining transparency about measurement procedures) by measuring within a time window, rather than at a specific date.

Without doubt, the contribution of Bevan and Hood (2006) to the field is outstanding. They analysed performance management in the English public health care system and found significant gaming and targeting problems. While they do not deny that there were strong improvements regarding the performance, they do not have certainty about whether these results are genuine or merely a trade-off at the expense of reduced performance in areas not measured by targets. Their recommendation is to include more uncertainty into the specifications of targets and how they are going to be measured, while being transparent about the process. This could help to counter targeting and gaming, they claim.

Contrary to Bevan and Hood's (2006) findings, Kelman and Friedman (2009) do not confirm the existence of gaming and targeting in the English National Health Service (NHS). They explain two potential dysfunctional effects, namely effort substitution and gaming. Effort substitution refers to the instance when where multiple goals are at stake, “what gets measured gets noticed” (Kelman & Friedman, 2009, p. 922). Consequently, improvement on one performance measure might be traded off by deterioration of results for another measure. The various forms of gaming have already been addressed here (Bevan & Hood, 2006; Smith, 1995). Kelman and Friedman (2009) then analyse econometrically all data of English public hospitals between 2003 and 2006 with regard to the so-called “four-hour wait time target”. This target implies that no patient in the emergency room should wait more than four hours for treatment. The authors find a dramatic increase in the performance improvement during the period under analysis, but no evidence for dysfunctional effects commonly associated with this target. Nevertheless, the authors admit that they were “unable to test the hypothesis about ambulances waiting outside A&E departments [i.e. ambulances waiting deliberately outside the emergency room department when the latter is too crowded to reach the four-hour wait time target] due to lack of available data.” (Kelman & Friedman, 2009, p. 926). This notwithstanding, they conclude that it is preferable to manage an organisation using imperfect performance measures than using none at all, if there is proof that the performance with measurement (and accompanying dysfunctional effects) is higher than without any measurement.

The difference between the results of Bevan and Hood (2006) and Kelman and Friedman (2009) can potentially be explained by examining their respective data collection methodologies. Kelman and Friedman (2009) conducted an econometric analysis of the data, which allowed them to draw conclusions on the officially recorded data. In contrast, by analysing additional studies and surveys other than those provided by the hospitals, Bevan and Hood (2006) accessed a different part of the story. This allowed them to dig deeper into the complexities of the performance management and target-setting system at stake. If this explanation holds valid, the important message to draw from it concerns the necessity for a close look behind the data, into the various mechanisms and interests at play.

In addition, a qualitative analysis by Hood (2006), in which the author analysed documentary sources and conducted interviews with high-level officials in the Whitehall bureaucracy, resulted in similar evidence for gaming. He found three known types of target gaming: the ratchet effect, the threshold effect, and output distortion or manipulation of reported results. Targets are set originally for performance improvement. However, they are seen to be “creatively interpreted, contrived, or spun”; or “invented, dropped, or [simply] not provided [at all]” (Hood, 2006, p. 519).

The argument that an imperfect performance measurement system is better than none at all, as claimed by – among others – Kelman and Friedman (2009), has received strong criticism, e.g. by Johansson (2015) who denounces the fact that despite mounting evidence of negative effects of performance management systems, the latter are rarely fundamentally questioned.

Even more pronouncedly, Radin (2006, 2000) accuses recent performance management reforms of attempting to fit “square pegs into round holes”. The complex issues of differences in context, programme structures, and goals of public actions run counter to a “one size fits all” approach as propagated by central performance management systems that focus mainly on efficiency (outcome) measures. The organisational realities seldom match the ambitions and goals of performance management systems.

Bohte and Meier (2000) analysed the extent to which school districts cheat to improve their students’ results in standardised exams. They find evidence for students with expected low pass rates being deliberately exempted from exams in order not to lower the overall result. Incentives for such cheating behaviour can be found in the fact that teachers’ salaries depend on their students’ results, and test results are used to rank schools and teachers. Scarce institutional resources and difficult task demands are seen to be the major triggers for organisational cheating. The authors hypothesise that cheating is more likely to occur in organisations that lack continuous monitoring by supervisors, and in which performance results are tied to incentive-based structures. To prevent cheating, Bohte and Meier (2000) suggest to focus on, and recognise, the important role of organisational norms against cheating.

Jung and Lee (2013) analyse the connection between organisational performance and job-goal commitment, mission specificity, and job-goal specificity in US federal government agencies. They find that the mentioned factors can all contribute to organisational performance. Interestingly, they note that in the public sector, intrinsic motivational factors play a more important role than extrinsic rewards, such as pay for performance. For example, trust among employees and between employees and managers, as well as effective communication skills, are highly valued, and as such, worth investing in. For organisational performance, this also means that perceptions and attitudes of employees regarding the organisation’s goals are important to take into account.

Nõmm and Randma-Liiv (2012) investigate why and how governments make performance information available, and how this information is used in decision-making. They examine these questions by means of a qualitative, exploratory analysis of the Estonian central government as an exemplary case of the introduction of performance management in post-communist countries. As the contextual factors are also influential in other central and eastern European countries, the authors argue that their findings can be generalised to a certain extent. Performance management was introduced in Estonia at the end of the 1990s – at the heat of the NPM movement in Western Europe, and thus also under the international demand for a performance measurement system. With a culture that emphasises the use of quantitative performance indicators, Estonia’s performance measurement system soon experienced an inflation of performance targets. However, at the same time, the administration lacks the necessary analytical, technical, and strategic capacities to use the system and the targets to their full potential.

Moving from the health care sector to the education sector, Nielsen (2014) analyses managerial authority and its effect on performance in Danish schools. He combines data

of standardised test scores of 45,000 students of 314 schools with a survey among school principals. He finds that managerial authority over human resources, including decentralised pay negotiations and hiring and firing employees, positively moderates the effect of performance management on organisational performance. Decentralised goal-setting, however, has a negative effect on performance management. The author concludes that even if no perverse effects such as gaming and cheating occur, performance measurement systems can be detrimental to organisational performance. A close look at contextual factors seems thus warranted when designing performance measurement systems.

To summarise, the previous sections have shown that evidence of dysfunctional effects of performance information is abundant. To bring some structure into the vast variety of effects, we can classify them according to two perspectives: first, the manipulation of measurement, and second, the manipulation of output (Van Dooren et al., 2015). Manipulation of measurement refers to “measurement that is not a good representation of reality”, but which does not affect “the real quantity, quality, and nature of the output of the organisation”. Manipulation of output, however, describes the dysfunctional effects of performance information, which “alter the daily operations of the organisation through the behaviour of organisational members”. This leads to “outputs of a different quantity, quality, or nature”. A third way constitutes the combination of measurement and output manipulation; manipulation of measurement can lead to manipulation of output, especially when the measurement system is flawed (Van Dooren et al., 2015, p. 184).

Manipulation of measurement can occur in several ways, including deliberate manipulation (such as misrepresentation) and “accidental” manipulation. The former describes what happened in Atlanta. Test scores were deliberately changed, which ultimately misrepresented the school’s and the students’ achievements, and in turn affected the teachers’ bonuses. The example described by Courty and Marschke (2007) also falls under this category; performance scores of job training programmes were increased by manipulating the measurement conditions. Accidental manipulation then includes over- and underrepresentation, “mushrooming” of indicator sets, and misinterpretation (Van Dooren et al., 2015).

Manipulation of output, on the other hand, influences the daily operations of an organisation or has an influence on the services provided. Measure fixation is one form of this kind of manipulation. It follows the “what’s measured gets done” logic, which results in either an oversupply of services or a decline in the quality of these services. The “teaching to the test” behaviour in public schools is an example of such an effect. Also De Koning's (2015) example of the Dutch police performance reports mentioned at the beginning of this chapter fall under this category; the focus on performance targets comes at the neglect of a focus on the bigger picture and qualitative services. In an earlier article, De Koning (2014) describes in greater detail the manipulations undertaken by the Dutch police, such as registering cases of housebreaking as garage or barn burglary, to make the housebreaking numbers appear to decrease. In addition, the “four-hour wait time” target of the English National Health Service qualifies as such

dysfunctional behaviour, which combines measurement and output manipulation, or the VA health centres' attempt to increase performance scores – with severe consequences. Other forms of output manipulation are “myopia,” cream skimming or cherry picking, and ratchet effects.

Box 1 below summarises the most prominent effects of the use of performance information, as presented in the present and the previous sections, distinguishing between manipulations of measurement and manipulations of output. The following section addresses the organisational and political contexts in which performance information plays a role.

Box 1 Effects of the use of performance information

Manipulation of measurement

- Indicator pollution
- Indicator inflation
- Mandelbrot disease
- Misrepresentation
- Misinterpretation
- Increased monitoring costs of the organisations and the state

Manipulation of output

- Tunnel vision
- Sub-optimisation
- Myopia
- Measure fixation
- Organisational paralysis
- Oversupply of outputs
- Lack of innovation and experimentation
- Symbolic behaviour
- Creaming
- Definition creep
- Lowering quality
- Encouragement of innovation in cost cutting, but not in service delivery
- Restrained overachievement
- No provision for start-up costs
- Reward of promises, not performance
- Focus on outputs instead of outcomes
- Undermining equity and fairness
- Ratchet effect
- Threshold effect

2.3. Organisational and political contexts

The list of negative consequences in the public sector could be extended easily. However, instead of repeating or enumerating all facets of unintended effects of performance measurement, let us examine the role of the context. Context is crucial when attempting to understand why and how these effects appear. This section thus focuses on the organisational and political contexts and their implications for the use of performance information.

Berman and Wang (2000) analyse the extent of the use of performance information and the necessary capacities for successful use in US counties with more than 50,000 inhabitants. They find that approximately a third of the counties under investigation use performance information, though to varying degrees. They find that technical capacity, i.e. the capacity to develop output and outcome indicators, and to collect, analyse, and evaluate the information strongly correlates with the use of performance information. In addition, stakeholder capacity plays an important role. While the support of county managers does not appear to make a difference between high and low users of performance information, support by heads of departments, supervisors, and elected officials does appear to make a great difference.

Noordegraaf and Abma (2003) investigate the management by measurement movement from a theoretical perspective. They analyse the pro- and con-debate on management by measurement, and develop a typology of management practices. They find that management by measurement is particularly suited for established, classified, and uncontested practices. Consequently, it is less appropriate for developing, changing, and highly contested practices. To gain a better understanding of (the appropriateness of) management by measurement in the public sector, it is indispensable to consider the context.

This is what Noordegraaf (2008) did in a study following the publication of the former article. He looked into the story behind the Safety Index of the Dutch city of Rotterdam, which numerically objectifies safety and “liveability” in Rotterdam. The author finds that to improve management by measurement, it is not so much necessary to improve methods and use of information. Rather, we have to examine the key players involved, how they act, and under what conditions.

In a recent publication, Kerpershoek et al. (2014) analyse the unintended responses to performance management in Dutch hospital care, by paying particular attention to the managerial and professional perspectives. They find several unintended effects, such as volume, diagnose, and treatment up-coding. However, they contend that to understand the complexity of the matter at hand, it is necessary to consider both financial and professional motives. They conclude that these unintended responses are not inevitably “perverse” as they are not only financially, but also professionally or value driven. Nevertheless, to counter the emergence of unintended effects, the authors suggest reducing the financial impact of the system and engaging in active dialogue between managers and professionals.

In a similar vein, Gregory and Lonti (2008) explore the use of performance information for policy advice in five New Zealand government departments. By analysing annual reports to Parliament between 1992 and 2005, the authors find that performance measures reflect what can easily be counted, rather than provide a valuable means for enriching policy advice. They attribute this also to the complexity and ambiguities of the public sector. Hence, Gregory and Lonti (2008) conclude that indicators are first and foremost important symbolically as opposed to practically.

Dahler-Larsen (2014) brings attention to the implicit notions of the terms “unintended” effects. Referring to Noordegraaf (2008), he explains that due to the complexities of the public sector and the social and political context in which measurement takes place, discussing “unintended” effects when dealing with (highly) contested issues might be somewhat naive. In fact, discussions about unintended effects assume that there are also “intended” effects. However, practice and theoretical literature fall short of evidence surrounding explicit intentional behavioural effects of performance measurement. Hence, Dahler-Larsen (2014) prefers to talk about constitutive effects, by which he refers to “something similar to a rhetorical effect, a meaning being constructed or a practice being established”, as opposed to causal effects. As such, indicators are constructed in organisational procedures, they provide a language as they name what is being measured, and they can also create “institutional lock-in” when indicators are linked to incentives, and, consequently, to sanctions in case of not achieving a certain score on an indicator. Thus, to understand the complexities and constitutive effects of indicators, context (such as working climate, managerial styles, publication, and media attention) and the role of stakeholders play an important part.

In sum, these contributions highlight the fact that the use of performance information (and the occurrence of negative, unintended, or “constitutive” effects) is strongly dependent on the context within which this information is used. Factors that play a role are the technical capacity to develop, collect, and analyse performance information, and stakeholder capacity, and particularly, the support by heads of departments or supervisors, but also by elected officials. However, not only is the support important, also the way these stakeholders interact with each other. Regarding the latter, it is suggested to invest more in active dialogue among the key players and, at the same time, reduce the tight coupling between financial means and performance information. However, it is not only financial motives that offer an incentive for gaming or cheating; often professional motives or values are also a driver for such behaviour. This is especially the case as the meaning of indicators is often more symbolical than practical. In particular, in the public sector, where practices are often contested and ambiguous, applying a “management by measurement” system is likely to have counterproductive effects.

2.4. Tension between internal improvement and external reporting

If we want to emphasise the recognition of underlying intentions of indicators when discussing unintended effects (Dahler-Larsen, 2014), this leads to another dilemma: the effects of performance information use when performance indicators are used for internal improvement and for external reporting. Several authors elaborate on this issue, of which a few are outlined below.

Nielsen and Baekgaard (2015) examine the impact of performance information on politicians' attitudes to spending and reform, in particular blame-avoidance strategies. By conducting a randomised survey experiment among 844 Danish city councillors, they determine that both high and low performance have a positive effect on attitudes to spending. However, if performance is reported to be average, this has a negative effect on spending attitudes. Willingness to reform is lower if performance scores are higher. The authors find that performance information does indeed affect the behaviour and the attitude of politicians. Blame avoidance seems to play an important role in politicians' responses to performance information. Hence, Nielsen and Baekgaard (2015) conclude that more attention should also be paid to high performance, rather than focusing on low performance only. This could lead to public officials shifting focus towards striving for success instead of focusing on avoiding blame.

Additionally, Hood (2007) examines the issue of blame avoidance, linking it to transparency. This is an interesting aspect to consider, as in the international arena efforts are undertaken to render governments and their policies more transparent. For example, as a member of the Open Government Partnership, the OECD plays an important role in supporting its member and candidate countries in "design[ing] and implement[ing] ... open government strategies" (OECD, 2015b). Hood (2007) departs from the negativity bias in relation to blame avoidance, as Nielsen and Baekgaard (2015) also explain. They contend that negative news or (policy) failures generally receive much more attention than stories of success do. Hood (2007) then explores theoretically the behaviour of elected and appointed officeholders at different levels in government and public services. He elaborates on three strategy types, namely, agency, presentational, and policy strategies. Agency strategies refer to blame shifting among the actors involved in the issues at stake. Presentational strategies denote the way outcomes are presented, or, in Hood's words, they are about "spin[ning] your way out of trouble" (p. 201). Policy strategies then describe the choice for low-risk routes and policies that are unlikely to create much blame in case of failure. Hood (2007) contends that transparency measures are likely to collide with blame avoidance strategies inherent in public policy. This, in turn, is likely to lead to futility, jeopardy, and perversity – terms which are framed by Hirschman (1991) for describing reactionary narratives. According to Hood, futility refers to

“outcomes of ‘dynamic conservatism’ in which no effective change occurs; jeopardy denotes outcomes in which effective change occurs, but at the cost of other

important values; perversity denotes reverse effect outcomes, in which the result is the opposite of that intended by reformers” (Hood, 2007, p. 203).

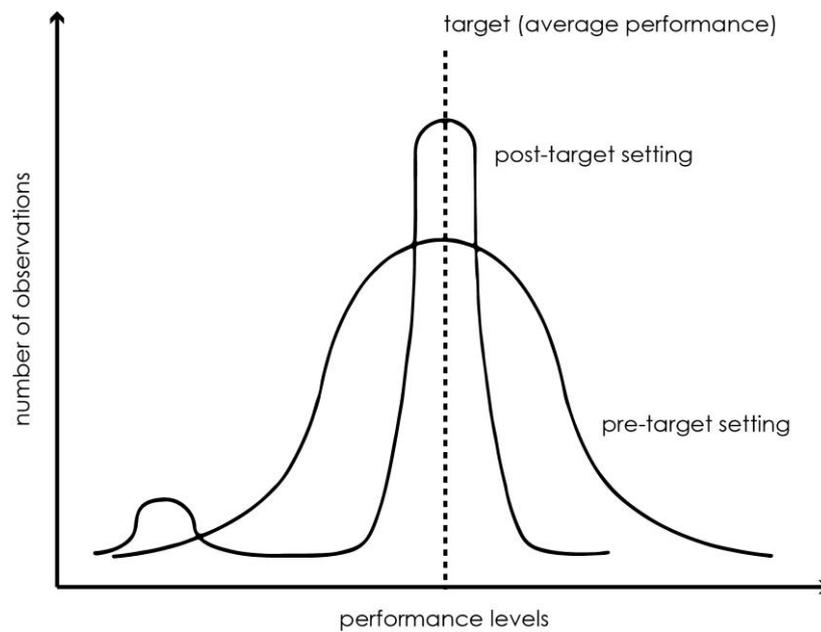
In addition, Wiggins and Tymms (2002) find that in English school systems, where key performance indicators play a central role, the level of “blame culture” (p. 46) is higher compared to Scottish schools, which do not have such indicators. This, in turn, can have negative consequences for the working environment, such as less collaboration and more competitiveness between teachers and schools.

In line with these findings, Alcock (2004) relates the occurrence of unintended effects to the high pressure of reporting for performance results. He examines public policy programmes in the UK focusing on combating poverty and social exclusion, and on improving the development and delivery of public services. He analysed documents, conducted interviews, held focus groups, and observed planning meetings to gain an in-depth view of the mechanisms at play. His results show that the perceived burden of reporting requirements led to a shift in targets from measuring outcomes to measuring inputs and outputs. Moreover, in particular in the area of social policy, the quantification of targets and reducing progress to simplified numbers led to a loss of measuring quality. This measurement culture resulted in practice being oriented towards meeting specific milestones and, as such, indicators became a target in themselves, rather than a means to reach that target. Consequently, this brought about managers setting softer, more easily reachable targets. Thus, unintended, perverse effects, as elaborated above, came about.

This is also what Van Dooren et al. (2015) call the “performance *target* paradox”, reflecting the work of Van Thiel and Leeuw (2002). The performance target paradox refers to the fact that target-setting influences behaviour in such a way that performance is concentrated around that target. This means that better performers will reduce their performance levels to avoid the ratchet effect mentioned earlier. Underperformers will also adjust their behaviour to reach the target, be it in a functional or dysfunctional way (see Figure 1). Support for the performance target paradox can also be found in the study of Wiggins and Tymms (2002) on league tables in English schools.

Alcock (2004) thus argues that it is mainly rigid timescales and budgets that place pressure on demonstrating evidence (and good performance). He shows that too strong a focus on monitoring for performance management can inhibit innovation and organisational development, as too much effort is put into activities with a backward-looking perspective rather than a forward-looking perspective, which could invoke the desired (structural) change.

Figure 1 Performance target paradox - the tendency towards a target



Source: Van Dooren et al. (2015, p. 192)

Similarly, Ammons (2002) elaborates on the contradiction between performance reporting and performance improvement, which according to him, is an important reason for the imbalance between promised and actual performance. He argues that this imbalance is because most performance measurement systems are designed for performance reporting, and, *per se*, accountability, whereas less (or no) emphasis is placed on performance improvement. Furthermore, returning to the negativity bias, small successes or improvements do not usually receive the necessary attention. Thirdly, performance measurement can be a catalyst for performance improvement, but it is not the source of performance improvement. As such, performance measurement seldom receives the credit it might indeed deserve. However, to exploit performance measurement more as a catalyst for performance improvement, Ammons (2002) suggests that old assumptions are challenged. For instance, he considers it important to start from the operational level and engage with the people there to determine their unit's or department's goals and roles in achieving improved performance. This would, ultimately, also be of use to the accountability needs of the higher levels. Nevertheless, this shift in approach would also mean that more emphasis should be placed on qualitative, instead of quantitative, measurement.

Last but not least, Goddard and Mannion (2004) investigate the advantages and disadvantages of horizontal, network, and vertical, hierarchical approaches to performance measurement and improvement in the UK public sector. They find that a predominantly top-down approach has seemingly (i.e. according to the performance indicators) led to improved performance. (The question, however, is whether performance has "genuinely" improved or whether the improvement is due to gaming). At the same time, there is abundant empirical evidence available that shows that this

approach is also likely to bring about unintended, dysfunctional effects. In particular, when sub-optimal performance or failures are placed in the spotlight by the media. Thus, the authors argue that a top-down approach is most likely to be successful when only a number of key targets are measured, which, in turn, are deemed important for central policy goals. A horizontal approach, on the other hand, appears to be useful for improving performance across the board and encouraging continuous learning, by also taking into account middle performers and front-line staff. Hence, effective ways to combine top-down, control, and accountability elements with bottom-up, networking, and learning elements have to be found to avoid dysfunctional effects and to make the most of the potential of performance information.

The contributions in this section show that a substantial tension exists between internal improvement and external reporting purposes. Based on the literature, this tension is due to the different underlying assumptions of these purposes; improvement is more forward looking and learning oriented, while reporting is more backward looking and control oriented. In turn, a particularly strong focus on control leads to blame avoidance strategies and a high pressure to perform well and report (positive) results. In addition, the negativity bias plays a role. Negative performance results receive far more attention from the media, citizens, civil servants, and politicians than do (small) successes. In fact, several authors explain that little to no attention is paid to successful performance results. As such, to avoid (public) blame, organisations adapt their behaviour accordingly through manipulations of measurement and outputs; setting easily achievable targets; and focussing on achieving the targets, while losing the focus on the bigger picture.

To gain a deeper understanding of this tension, Van Dooren et al. (2015) offer a distinction between soft and hard use of performance information. If performance measurement directly leads to performance assessment, we speak of so-called tight coupling (Van Dooren et al., 2015, p. 122). Tight coupling generally does not leave room for the consideration of the context before a final assessment is made. Loose coupling, on the other hand, refers to exactly the opposite. In fact, dialogue, interpretation, and consideration of the context mediate the judgement based on performance measurement. In particular, when performance results lead to direct consequences (e.g. budget cuts in case of under-performance), performance judgement has a high impact on an organisation or the employees, and performance information is thus used in a hard way. Thus, we can expect more “dysfunctional behavioural effects” (Van Dooren et al., 2015, p. 124) in this scenario than in the case of soft use.

Nevertheless, what might be considered hard use in one specific context might be less so in a different context. Van Dooren et al. (2015) mention the example of school league tables. While these league tables play a major role in, for example, British parents' decisions regarding the school to which they will send their children, in other countries other factors are taken into account. As such, league tables can be considered an example of hard use (performance information results lead to direct judgement) in one country's context, whereas they are just another source of information (and as such, used in a softer way) in a different country's context.

2.5. Conclusion and implications

This chapter focused on the effects of the use of performance information. It has become apparent that the use of performance information affects the users' behaviour. Effects can be distinguished according to manipulations of measurement and manipulations of output, with "manipulation" referring both to deliberate and unintended behavioural changes. While the former does not necessarily affect the behaviour of the members of the organisation, the latter does so. This leads to a change in the service provision, and as such, the consequences of output manipulations are significant.

The literature review has shown that evidence for negative effects of performance information is far from anecdotal. This is particularly the case when performance information is used in a "hard" way, i.e. when performance information is tightly coupled to judgement and consequences, as would be the case in performance contracting, pay for performance systems, school rankings, etc. Tightly coupled systems are generally associated with accountability and control purposes, with the overall aim of making the system more efficient (Van Dooren et al., 2015).

On the other hand, several authors have suggested to move beyond the control-based systems towards learning-based systems, or to merge these two approaches. Learning-based systems would be based on engaging with stakeholders, communicating clearly and establishing dialogue among managers and employees (see also Fryer, Antony & Ogden (2009) in this regard), keeping the performance measurement system flexible and dynamic, and focusing on qualitative measures in addition to quantitative ones. This approach would reflect a loosely coupled system, where judgement and consequences are mediated by dialogue.

Nevertheless, empirical evidence for learning-based systems is relatively scarce, in comparison to control- or accountability-based systems (exceptions are, e.g. Moynihan, 2005 and Olejniczak & Mazur, 2014). Nor is there abundant evidence for the co-existence of these two systems. This is not too surprising, as these two views are contradictory at first sight. After all, a completely different vision of the performance measurement system, its purpose, and its use is disclosed.

It is at this point from which this dissertation departs. Further (empirical) evidence is needed for the use of performance information in learning-based performance management systems, and the compatibility of using this information in accountability- and learning-based systems.

To provide the empirical part of this dissertation with a conceptual and analytical framework, the next chapter examines more closely the actual use of performance information, as already alluded to in some of the articles reviewed for the present chapter.

3. USE OF PERFORMANCE INFORMATION

*How you gather, manage and use information
will determine whether you win or lose.
~ Bill Gates*

Performance information is supposed to play a vital role in the daily management of an organisation and should inform, among others, personnel decisions, strategic decisions, advocacy purposes, the allocation of resources, and efficiency purposes (Moynihan & Hawes, 2012). At the same time, we have seen in the previous chapter that using performance information for these purposes can have unintended or even dysfunctional effects. To understand where these effects might originate from, it is important to examine how performance information contributes to decision-making. This chapter analyses the different scholarly approaches to the use and purpose of performance information.

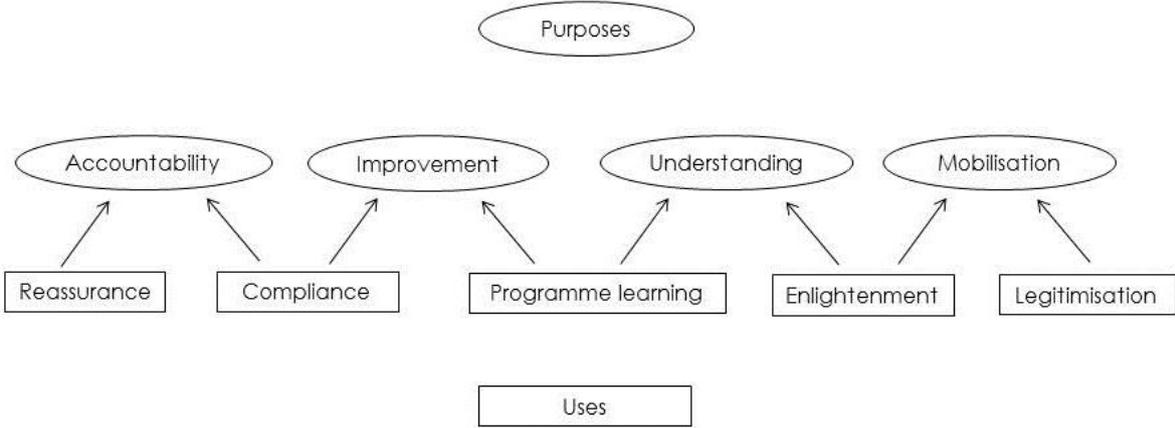
3.1. The different uses and purposes of performance information

A first important point is raised by de Lancer Julnes (2008). She emphasises the distinction between purposes and uses of performance information, which require different kinds of performance measurement information. Purpose refers to “knowledge converted into action or implementation of performance measures”, while use describes the “underlying concept that provides meaning to purposes” (de Lancer Julnes, 2008, p. 67). Her typology consists of four purposes, i.e. accountability, improvement, understanding, and mobilisation. However, she claims that in practice, performance information is rarely used for improvement purposes. These purposes correspond to five types of uses, namely reassurance, compliance, programme learning, enlightenment, and legitimisation (see Figure 2).

De Lancer Julnes (2008) contends that for performance information to be used as intended, both rational factors, i.e. the technical capacity of an organisation to use information, and political factors are necessary. Technical capacities are necessary to “adopt” performance information, i.e. to develop the performance management system, and as such, the “capacity to act”. These include internal support, availability of resources, technical knowledge, and goal orientation (de Lancer Julnes, 2008, p. 64). However, to implement this information, political factors are also necessary. Among these are leadership, the “primacy of politics in performance measurement”, political ideology, and partisanship. As a result, there is no “one size fits all” model of performance measurement. “Different audiences, different purposes, and different uses require different kinds of performance measurement information ... Consequently, the

system has to be flexible enough to adjust to different requirements for information. In essence, the system has to learn.” (de Lancer Julnes, 2008, p. 75). The need for flexibility has already been discussed in the previous chapter (e.g. Smith, 1995). The learning capacity is addressed in section 3.3.

Figure 2 Types of knowledge use and purpose



Source: de Lancer Julnes (2008, p. 71)

Similarly, Behn (2003) suggests a list of eight managerial uses of performance information (see Figure 3), although he uses the concepts of use and purpose interchangeably. He concludes that each of these uses requires a different baseline against which performance is to be compared. This baseline, in turn, is highly dependent on the context.

Figure 3 Eight purposes of use of performance information

-
- | | |
|------------|-------------|
| - Evaluate | - Promote |
| - Control | - Celebrate |
| - Budget | - Learn |
| - Motivate | - Improve |
-

Source: cited from Behn (2003)

Another, more abstract typology of uses is provided by Moynihan (2009). He distinguishes between purposeful, passive, political, and perverse use.

Purposeful use refers to what de Lancer Julnes (2008) describes as programme learning, or the purpose of improvement. Performance information is used to improve governance. This can be done by means of “hard use”, such as applying sanctions and rewards, as well as by means of “soft use”, i.e. as a tool to support learning and better decision-making (more on the distinction between soft and hard use below). Purposeful use is in

most cases the type of use that is considered when discussing the results of performance management or the effects of performance information; it is assumed that performance information has a certain intention or purpose. Against this implicit purpose its success is assessed (Moynihan, 2009).

Passive use is about complying with the creation and dissemination of data, but not using this data. “Ticking the box” exercises can be a consequence of such use, or rather, non-use. Nevertheless, Van Dooren et al. (2015) argue that there might be a variety of reasons for non-use of performance information. For example, information could be used in informal, undocumented ways, and as such, use is not “visible” or “measurable”. However, the authors also provide some possible explanations for the non-use, such as insufficient quality of the information, psychological barriers, cultural barriers, and institutional contexts.

Political use refers to using information as a “means of advocacy”. Actors select and interpret information according to their own needs, and to underline their efforts and successes. This can be linked to de Lancer Julnes’ use for legitimisation.

Finally, *perverse use* of performance information is likely to occur in situations when actors are under pressure to perform well. The previous chapter has shown that under these conditions of perceived “hard use” of performance information, cheating, gaming, and setting up sub-optimal indicators, which in the end might even contradict the original goal, are the consequence (Moynihan, 2009).

Another list of different uses is suggested by Hatry (2011). He proposes a taxonomy of eight different uses, which he summarises in three main categories of use: Accountability, budgeting, and improving services. The latter includes several sub-uses, among them performance contracting, providing information for in-depth evaluations, learning/problem-solving, motivating employees, communication with citizens, and marketing/advocacy. Thus, they also reflect de Lancer Julnes' (2008) typology.

To categorise the variety of uses and purposes in broader terms, Van Dooren et al. (2015) distinguish between learning, steering and controlling, and giving account as three major purposes of use (see Table 1). Learning is about improving policy or management; steering and controlling aims at finding ways to control activities; and giving account is about the communication of the performance. Consequently, learning is oriented towards change and the future, steering and controlling focuses on the present, and giving account has a backward-looking character.

Table 1 Three uses of performance information

	To learn	To steer & control	To give account
Key question	How to improve policy or management?	How to be in control of activities?	How to communicate performance?
Orientation	Change / future	Control / present	Survival / past

Source: Van Dooren et al. (2015, p. 120)

Similarly, Olejniczak and Newcomer (2014) distinguish between performance accountability and learning (see Table 2). While performance accountability is driven by following procedures and fulfilling plans without errors, learning emphasises experiments and the use of trial and error processes. Consequently, accountability focuses on processes and clearly measurable products. Learning, on the other hand, looks at desired outcomes and analyses trends in performance over time. This leads to accountability repeating what works, i.e. exploitative behaviour, while learning stresses exploratory behaviour by searching for new solutions. Another distinction is the performance information used in these two different approaches; accountability systems rely on routine information, structured data, and the monitoring of indicators. Learning, on the other hand, resorts to both routine and non-routine information, and opinions and feedback from diversified sources.

Table 2 Performance accountability and organisational learning compared

	Performance accountability	Organisational learning
Driving logic	Follow procedures and fulfil the plans without errors	Experiment, use trial and error processes
Scope of performance	Focus on process and clearly measurable products, here and now	Envision desired outcomes and analyse trends in performance over time
Information basis	Routine information, structured data, monitoring and indicators	Non-routine information, opinions, feedback from diversified sources
Typical behaviour	Exploitation – repeat what already works	Exploration – search for new solutions

Source: Olejniczak & Newcomer (2014, p. 85)

When we now examine the conceptualisations of Van Dooren et al. (2015) and Olejniczak and Newcomer (2014), we see that what the latter call “performance accountability”, comprises the concepts of steering and control and giving account, as posited by Van Dooren et al. (2015). In fact, as shown in the following sections, accountability cannot be thought of without considering control mechanisms. Furthermore, when we consider the

academic literature, accountability and learning are the types of use of performance information that receive most attention. In addition, as their underlying expectations and behaviours are – in their “pure” forms – considerably distinct, it makes sense to focus attention on these two forms of use. In addition, the previous chapter has shown that the effects of the use of performance information can be quite remarkable – to the extent that they turn all performance measurement efforts even into perversity. The following subsections examine more closely the elements of accountability and respectively learning systems, and the (in)compatibility of these two.

3.2. Accountability in performance management systems

Accountability is one of these “magic concepts” in the sense that it is “very broad, normatively charged and lay[s] claim to universal or near universal-application.” (Pollitt & Hupe, 2011, p. 643). Ostensibly, it is difficult to find reasons for opposing such a concept. Nevertheless, recently, also “the other side” of accountability has been examined. While some claim that accountability is “one of those golden concepts that no one can be against” (Bovens, Schillemans & 't Hart, 2008, p. 225), others contend that, particularly from the accountability givers’ point of view, “accountability means punishment” (Behn, 2001, p. 3). Thus, what does accountability actually mean?

The first sub-section elaborates on the conceptualisation of accountability and the link between accountability and performance. The second one examines more closely the so-called accountability paradox, before the third sub-section presents the complicated relation between accountability and trust, as already alluded to in the first chapter.

3.2.1 Accountability – an approximation to the concept

Accountability can be defined as “a relationship between an accountant and an account-holder in a forum, in which the accountant has an obligation to explain and to justify his or her conduct, the accountholder can pose questions and pass judgement, and the accountant may face consequences” (Willems & Van Dooren, 2012, p. 1018).

This definition reveals three main components of accountability: Information, debate, and judgement and consequences (Bovens, 2007; Schillemans, 2008; Willems & Van Dooren, 2011, 2012). These components play out in different forums. Based on the literature on accountability, Van Dooren and Willems (forthcoming), Willems and Van Dooren (2012), and Mulgan (2003) identify five such forums, which include the administrative, the political, the judicial, the public, and the market forums. In each of these forums, actors are held to account through different processes (see Table 3). Importantly, actors are not required to justify their actions in one single forum only. Rather, they can be held accountable in various forums, at any moment, at the same time.

Table 3 Forums and processes of accountability

Accountability forums	Processes of accountability
Political	Elections, parliamentary scrutiny, political debate within political parties, ministerial responsibility, ...
Administrative	Hierarchical command, diverse government auditors, regulatory bodies, ombudsmen, other types of government monitoring agencies, performance contracts, budgets, ...
Judicial	Judicial courts, administrative tribunals, mediation ...
Public	Mass media, social media, civic action, ...
Market	Shareholders' reports, consumer choice, rating agencies, ...

Source: Van Dooren & Willems (forthcoming)

In addition to accountability forums, accountability functions or purposes have to be considered able to understand the complexity of the concept. In line with Bovens et al. (2008), Willems and Van Dooren (2012) suggest three accountability functions: Constitutional, democratic, and performance. Bovens et al. (2008) call the latter the “learning perspective”. This already shows the theoretical link between accountability and learning.) The constitutional function refers to governments being held accountable for following the “rules and procedures to prevent unfairness or abuse of power”. The democratic function describes the power and right of citizens to hold their governments to account. The performance function touches on accountability for results – the answer to the question, “what government actually accomplishes” rather than “how” (Willems & Van Dooren, 2012, p. 1023). It also draws attention to the learning capacity of governments, which ultimately is to improve government effectiveness (see also Aucoin & Heintzman, 2000; Bovens et al., 2008; Willems & Van Dooren, 2011). Accountability for performance is of particular interest here. This perspective is about providing the right incentives and high quality feedback, and engaging in debate and dialogue. However, while these are laudable purposes of accountability, according to Bovens et al. (2008), they are difficult to achieve in reality.

This conceptualisation shows that accountability takes place in different forums with different purposes. Accountability can be a hierarchical form of control, when account-holders are superior to those being held accountable. Think of employees being accountable to their managers, managers of a government agency being accountable to their line ministers, etc. However, accountability also works horizontally, when account-holders are not superior. The media, peers, or stakeholders are such non-superior account-holders (Schillemans, 2008). Horizontal accountability is considered a “complementary”, not a “substitutional” form of vertical accountability. Its advantage lies in its potential to foster dialogue and learning. However, the power of horizontal accountability should not be under-estimated either. Horizontal accountability lives of “social control”. However, just as in situations of vertical accountability, nobody wants to lose reputation or legitimacy in horizontal accountability arrangements. Thus, the risk of

“defensive routines” is still present and not un-justified, albeit to a lesser extent than in vertical accountability arrangements (Schillemans, 2008). In addition, all these “controlling actors” have different expectations. As such, accountability becomes “a balancing act among multiple sources of control” (Willems & Van Dooren, 2011).

Against this background, let us examine the dysfunctional effects of performance information again. We have seen that they are often caused by too much a focus on achieving set goals and milestones (i.e., too much a focus on meeting different expectations and avoiding blame at the same time). Why do these effects appear? Which mechanisms play out when governments are held accountable for their performance, based on performance indicators?

Performance- or results-based management has become the centre of many public administration reforms in the last decades (Moynihan, 2008; Pollitt & Bouckaert, 2011; Van Dooren et al., 2015). Accountability for performance (or, more specifically, indicator-based accountability) occurs, thus, for example, through annual reports, contracts, or performance budgets. Hence, it takes place primarily in the administrative forum. Public managers and civil servants are held to account for their results by their superiors within the administration and/or executive politicians. Nevertheless, the story does not end here. Performance indicators seldom remain within the “boundaries” of the administrative forum. Rather, a “spill-over” into other forums is likely. Actors are therefore held accountable for their actions also in forums other than the administrative one. (Moreover, in forums the indicators might not have been intended for.) This is what Willems and Van Dooren (2012) term “multiple accountabilities”. Take, for example, the political forum. Literature finds that generally, performance information is not much used by politicians (Currstine, 2006; Sterck, 2007; Talbot & Talbot, 2008; ter Bogt, 2004). Nevertheless, occasionally, such performance information does enter the political arena. Also in the public forum can performance indicators make their appearance, in particular when targets are not met. The negativity bias has been mentioned already in the previous chapter (Hood, 2007; Nielsen & Baekgaard, 2015). In addition, under certain circumstances, a spill over into the market forum can also occur.

The Greek crisis is a prime example. Without going too much into detail about the causes and drivers of the crisis in general, major problems were found within the Greek public administration. The latter is now held to account in not only the administrative, but also the political, and the public forums. Regarding the spillover to the market forum, just look at the reaction of rating agencies and the stock markets to the situation in Greece.

To keep it simple, accountability is complex. Morieux (2015) claims that “clarity, accountability, measurement were OK when the world was simpler.” At the same time, there is little contestation about the necessity of accountability. So which way to go forward? The next section outlines the discussion on the “accountability paradox”.

3.2.2 The accountability paradox

The multidimensionality of accountability has consequences. In fact, the introduction of performance-based management followed the doctrine of “doing more with less”, making bureaucracy more efficient and effective with less means. Ultimately, these three Es of efficiency, effectiveness, and economy ought to lead to an improvement of performance (Osborne & Gaebler, 1992). Nevertheless, as the previous chapter shows, empirical evidence for such an improvement is scarce. Evidence for gaming and cheating, however, is abundant. Interestingly, even in the private sector, the “basic tenets about efficiency and effectiveness have become counterproductive for human efforts”, Morieux (2015) condemns, and raises the question whether the “holy trinity of efficiency, [i.e.] clarity – measurement – accountability” is in danger. Just think of the VW emissions scandal that became known in autumn 2015 – management by results does not work problem-free in the private sector, either.

Indeed, Dubnick (2005) contends that we are confronted with an “accountability paradox”, and while the general assumption is that more accountability leads to improved performance, it might be just the opposite; the more accountability, the less improved performance. Alternatively, framed differently, accountability can have a detrimental effect on performance. Dubnick (2005) elaborates on the often-assumed relationship between performance and accountability. He does not find any empirical or theoretical evidence for this relationship, and concludes that to understand the relationship between accountability and performance, other context factors have to be taken into account, in particular expectations (from both sides, the account-holder and the accountor) and trust.

But also Halachmi (2002) claims that measurement for accountability and measurement for performance improvement are likely to be incompatible, due to the different natures of the concepts of performance and accountability. He asserts that accountability is about the relationship between the account-holder and the account-giver, as well as about the content of that for which someone is held accountable. Performance, on the other hand, is about progress, innovation, and change. This also implies creativity and adaptation, and out of the box thinking. Accountability, in contrast, encourages staying within in the box, i.e. the “contract” between the account-holder and the account-giver.

Furthermore, he contends that performance measurement for accountability can even undermine efforts to improve performance. Performance measurement for accountability, so Halachmi (2002, pp. 371–372), asks “was it done right?”, while performance measurement for improved performance is about the question “was the right thing done?”. Thus, accountability is about verifying past actions, mostly related to the proper use of resources, while improved performance is about learning from experience and exploration. This distinction links back to the previous section, in which the difference between the various types of purposes of performance information is addressed.

In line with this, Halachmi (2002) elaborates on the distinction between strategic planning and strategic thinking, and argues that strategic planning and accountability

are aligned, while this is less so, or even contradictory, in the case of strategic thinking and accountability. Strategic thinking is about thinking outside the box, being creative, and using one's intuition. Strategic planning on the other hand is based on a rational mind-set, relies on tacit knowledge, and follows established rules and norms. Linking this to performance measurement now, Halachmi concludes that when performance is measured for accountability purposes, this cannot lead to improved performance or out of the box thinking, due to the inherent characteristics of accountability measurement and strategic planning. Furthermore, he holds that in an organisation employees are likely to be dis-incentivised to go off the beaten track and discover new routes if they expect to encounter "negative rewards". In a similar vein, Flinders (2011) elaborates on the politics of accountability. He contends that there is a strong focus on the allocation of blame and negative results. This offers little incentives for politicians or civil servants to become involved in accountability processes.

At the same time, Halachmi's (2002) argumentation also works in the opposite direction. Measurement for improved performance is unlikely to be beneficial for accountability, as the time horizons for goal achievement (short term vs. long term) as well as the nature of the results differ. Thus, while both measurements are legitimate, using the same data for accountability and improved performance measurements is improbable to play out as intended (see also, e.g. Ebrahim, 2005).

If we link these findings back to the uses of performance information for accountability and learning, an important lesson from the literature on accountability is the following: While in theory performance indicators can be perfectly used for learning purposes, the "intermediary" factors are too unpredictable for this to happen. Thus, it is likely that indicator-based accountability encourages gaming or cheating.

Similar to Halachmi (2002), Van Dooren and Willems (forthcoming) claim that performance information for accountability purposes may contradict innovation and risk taking in organisations. This is due to several factors:

- Reputational damage
- Pressure in quasi markets
- Political blame game / recuperation
- Process instead of output / outcome
- Ratchet effect
- S-shaped performance improvement curve
- Unpredictability, which leads to uncertainty and risk avoidance. This, in turn, counters incentives for experimentation.

In summary, the accountability paradox implies that an increase in accountability does not necessarily lead to improved performance. Findings from various scholars support this claim by demonstrating that the underlying premises of accountability and performance improvement contradict each other. Accountability is more backward looking, while performance improvement is more forward-looking. Accountability

encourages a (strict) focus on reaching the set targets, learning encourages out of the box thinking and experimentation.

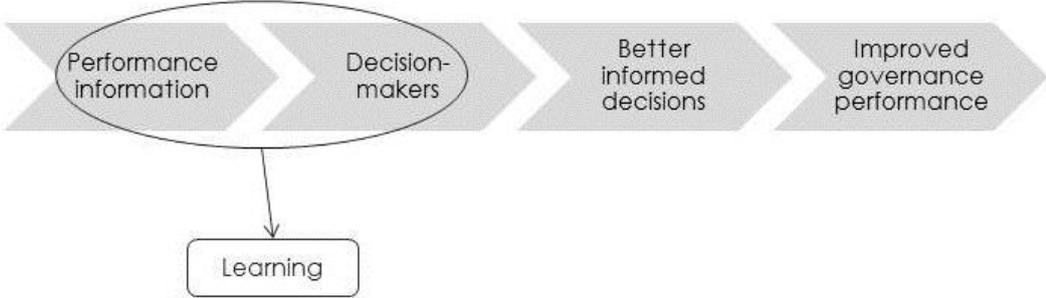
3.3. Learning in performance management systems

Similar to accountability, learning is a concept few could oppose. The Oxford Dictionary defines learning as “the acquisition of knowledge or skills through study, experience, or being taught.” Senge (2006, p. 13) claims that “[r]eal learning gets to the heart of what it means to be human”.

Nevertheless, if we elevate the concept of learning to a systems level, also this concept has received criticism. For example, learning systems are just a new form of control (Coopey, 1995), they are just a fashion (Furnham, 2004; Glaser, 1997), or they might provide some theoretical insights, but lack practical applicability due to their vagueness (Caldwell, 2012; Grieves, 2008). Furthermore, rather than substituting accountability efforts, learning systems would complement them, and as such increase the administrative burden on the organisation and create an information overload (Lewis & Triantafillou, 2012).

This notwithstanding, moving towards “learning systems” has increasingly been propagated. For example, Moynihan (2008, p. 164) claims that the performance management doctrine is essentially based on a “theory of learning”; but [has been] weak in identifying routes to learning.” He describes recent efforts of performance management as essential “organisational learning mechanisms” in that they assume that “decision makers will learn from performance information, and, in turn, they will make better-informed decisions and improve government performance” (Moynihan, 2005, p. 203) (see section 3.3.4). Organisational learning thus describes the “use [of] information and experience to make better decisions” (Moynihan & Landuyt, 2009, p. 1097).

Figure 4 Organisational learning and use of performance information



Source: based on Moynihan (2005)

This section with its subsections tries to shed light on the concepts of organisational learning and the learning organisation, as well as types of organisational learning. It also pays attention to the situational context in which learning is embedded. The last subsection recapitulates the insights from the present and the previous chapters to frame the theoretical context against which the empirical research following thereafter is situated.

3.3.1 Organisational learning – an approximation to the concept

Organisational learning is not a new concept in management studies. An influential – maybe the *most* influential– contribution to the academic literature dates back to 1978 (Argyris & Schön, 1978). In this contribution, the authors elaborate on the concept of organisational learning and the different forms learning can take.

However, organisational learning is a concept which has been developed and applied primarily to the private sector (for a review see Easterby-Smith, Crossan & Nicolini, 2000). It is only recently that organisational learning has found its way into the studies of the public sector as well.

Let us start with a few definitions of organisational learning. Organisational learning is

“when members of the organisation act as learning agents for the organisation, responding to changes in the internal and external environments of the organisation by detecting and correcting errors in organisational theory-in-use, and embedding the results of their inquiry in private images and shared maps of organization.” (Argyris & Schön, 1978, p. 29).

“the development of structures and procedures that improve the problem-solving capacity of an organisation and make it better prepared for the future.” (Olsen & Peters, 1996).

“the ability of an organisation to demonstrate that it can learn collectively by applying new knowledge to the policy process or innovation in policy implementation.” (Common, 2004).

“the adaptation that is based on the social process of reflection that produces new insights, knowledge and association between past actions, the effectiveness of those actions and future actions.” (Olejniczak, 2014, p. 14).

What do these definitions have in common? First, they focus on gaining knowledge and consequently using it for future actions. Second, they stress that organisational learning is, in essence, a process, as opposed to a static construct. Third, they emphasise the structural and cultural embeddedness.

Let us examine a few contributions to the general idea of organisational learning. Easterby-Smith et al. (2000) present the key debates of the past, present, and future about organisational learning, mainly from a private sector perspective. These debates include the distinction between individual and organisational learning, single- versus double-loop learning, and the distinction between organisational learning and the learning organisation. Remarkable, regarding the latter, the authors conclude that this

distinction is unlikely to play a role in future debates, as these concepts continue to converge. While this reasoning is understandable, I do see a conceptual difference between organisational learning and the learning organisation, which I discuss below.

Easterby-Smith et al. (2000) close their literature review with referring to three debates, that are likely to play a role in the future: 1) A regard for practice and activity as new units of analysis, i.e. an interest in how learning happens in the work place; 2) a focus on variety and difference as foundation and result of learning processes, i.e. the interest in learning processes, especially when different kinds of knowledge are involved, and where dialogue also plays a crucial role; and 3) attention for issues of power, politics, and control; i.e. among others, the establishment of learning forums, a concept I discuss further below, and an interest in “conflicts between shareholders’ goals, economic pressure, institutionalized professional interests and political agendas”.

Gilson, Dunleavy and Tinkler (2009) and Rashman, Withers and Hartley (2009) provide two recent reviews of the literature on organisational learning in the public sector, with important contributions dating back several decades ago (Argyris & Schön, 1978). Both groups of researchers find the concept of organisational learning to be underdeveloped in the public sector in comparison to studies of and applications to the private sector. Gilson and colleagues approach the organisational learning literature from a theoretical perspective and based on the experiences in the UK. They start with the definitions of the concept and the debates surrounding it, continue with the characteristics of organisational learning in the public sector, and finally address the wider context of influences, such as innovation, human resources management, political processes, and sources of organisational learning in governments.

Rashman et al. (2009) also include the concept of “organisational knowledge” in their systematic literature review, as well as the processes of inter-organisational learning, the relationships between source and recipient organisations, and the outcomes and measures of organisational learning. They also acknowledge the role of structure and culture in determining the success of organisational learning efforts.

In the first chapter of their edited “Handbook of organisational learning and knowledge management” (Easterby-Smith & Lyles, 2011a), Easterby-Smith and Lyles (2011b, p. 3) provide some general definitions of the concepts of organisational learning and the learning organisation. They define organisational learning as “the study of the learning processes of and within organisations, largely from an academic point of view” with the aim “to understand and critique what is taking place.”

Another important aspect is the relationship between individual and organisational learning. Is organisational learning “just” the sum of individuals’ learning? Is it more? Is it less? Hedberg (1981, p. 6) offers the following clarification:

“Although organizational learning occurs through individuals, it would be a mistake to conclude that organizational learning is nothing but the cumulative result of their members’ learning. Members come and go, and leadership changes, but organizations’ memories preserve certain behaviours, mental maps, norms, and values over time”.

Argyris and Schön (1978) explain that organisations are more than a collection of individuals. At the same time, organisations only exist due to such collections. The same may be said about learning; organisational learning is not simply the sum of individual learning, but without individual experiences and actions, organisations cannot learn. Framed differently, “there is no organisational learning without individual learning, and ... individual learning is a necessary but insufficient condition for organizational learning”. “Collaborative inquiry of individual members” serves as mediator to achieve organisational learning (Argyris & Schön, 1978, p. 20).

In brief, organisational learning describes the process of gaining knowledge in an organisation. This process is embedded in and fostered by structural and cultural factors, and is discussed further below. Organisational learning is not simply the sum of individual learning – it can be more, but it can also be less.

However, how does organisational learning manifest itself? Generally, the literature distinguishes between two main forms, which are examined in the following section.

3.3.2 Types of organisational learning: Single-loop vs. double-loop learning

The organisational learning literature distinguishes two main forms of learning, i.e. single-loop and double-loop learning (Argyris & Schön, 1978). Whereas in the context of performance measurement single-loop learning describes the attempts to improve the current, existing measurement system by asking how to do the same things better, double-loop learning questions the measurement system itself. It furthermore seeks to determine ways to measure differently (Common, 2004; Easterby-Smith, Crossan & Nicolini, 2000; Gilson et al., 2009; Moynihan, 2005, Moynihan, 2008, 2011; Rogers & Williams, 2006; Van Dooren, 2011).

Argyris and Schön (1978) describe learning as an “error-detection-and-correction process”. Single-loop learning is the result of error detection and correction, which makes organisations continue with their policies and achieve their objectives. Double-loop learning goes a step further. It implies correcting errors by questioning (and changing) norms, policies, and objectives of an organisation. Consequently, single-loop learning is easier to achieve than double-loop learning.

Recall the claim of Al-Ani (2015) (quoted in the introduction): Rather than focusing on new procedures and innovative technologies (i.e. investment in double-loop learning, how can things be done differently), organisations stick to the “known strategies, which are more often than not insufficient” – i.e. a form of single-loop learning.

In fact, Argyris and Schön (1978) argue that most organisations focus more on single-loop than on double-loop learning. Furthermore, they contend that organisations also occasionally deliberately opt for learning systems that inhibit double-loop learning. As double-loop learning is not only about detecting errors, but also correcting them in a way that also questions underlying norms or structures, this is often related to “conflicts of interests”, either between different members or requirements of an organisation.

As such, and this is another point raised by Argyris & Schön (1978), learning is not always necessarily a “good thing”. While it is common sense to associate “learning” with something positive, with progress and positive change, it can easily also be the opposite. Examine the discussion on the effects of using performance indicators for accountability purposes: Organisations and members of the organisations have learned how to manipulate data and measurement, how to achieve better performance scores – without changing their policies. This is also “learning”, albeit a different kind of learning. Similarly for accountability, it can thus be posited that for “positive learning” to occur, the organisational culture and environment also have to be favourable towards it. This is discussed more in detail in section 3.3.4.

With reference to the distinction between single- and double-loop learning, Easterby-Smith et al. (2000) contend that this division is somewhat blurred, as what might look like single-loop learning (or “incremental learning”) on individual level, might in fact be considered “transformational”, thus, double-loop, learning on the organisational level. Nevertheless, they see a shift of focus on double-loop learning. But also Argyris and Schön (1978) claim that the distinction between single- and double-loop learning is less black and white than it might seem. They see it rather as a continuum. As such, we can even speak of “more or less double-loop learning”.

Common (2004), however, argues that due to the “bureaucratic characteristics and strong departmental cultures” double-loop learning is not likely to occur in the public sector, but rather in the political sphere. According to him, single-loop learning seems thus the preferred mode of learning in the public sector. Similarly, also Moynihan (2008) contends that the focus of performance management has primarily been on single-loop learning. Goals are rarely challenged, and priority is given to finding ways to improve the achievement of these goals. Nor are public administrators encouraged to challenge these goals.

Finally, there is also *deutero*, or triple-loop, learning (Argyris & Schön, 1978; Gilson et al., 2009; Rogers & Williams, 2006). This form of learning describes efforts to “learn to learn”. It is about reflecting on current practices, norms, assumptions, and beliefs. As such, it goes a step further than double-loop learning, and can be seen as a logical, albeit challenging, continuation of double-loop (but also single-loop) learning efforts.

Let us use the example of performance indicators to illustrate the different forms learning can take. First, single-loop learning: Performance indicators are set to reach a specific outcome. Single-loop learning asks, was it done right? Thus, it is about detecting errors and correcting them to reach the aspired outcome. The performance indicator on which the outcome measurement is based is not questioned, and focus lies on achieving the outcome more efficiently. Van Dooren (2011) suggests four strategies for single-loop learning which essentially amount to a better implementation of the (existing) performance management system. These strategies include improving the quality of performance information, considering (and acting upon) the important role of leadership, investing in ownership, and integration and coordination. Nevertheless, all these strategies come with their own costs, limitations, and potential pitfalls.

Double-loop learning offers a different way of looking at the improvement of performance management systems. Double-loop learning asks, was the right thing done? Hence, here we discuss finding ways of doing it differently, for example, considering changing the performance indicators. In a way, this is what has happened in recent performance management reforms, where focus shifted from measuring performance by means of input and process, to output and outcome indicators. However, double-loop learning also includes the broader question of whether things can be done differently, for example, by questioning the performance measurement system as such.

The main strategy put forward by Van Dooren (2011) concerns the ambiguity of performance management systems. Rather than denying or ignoring ambiguity, we should recognise that performance management is complex and ambiguous. Consequently, performance management systems must be ambiguity-proof. This includes making sense of performance information by enabling a performance dialogue, which takes into account various sources of information, situational and context knowledge, values, politics, and conflicts (see in this context also Moynihan's (2008) elaborations; further explained in section 3.3.4). In turn, this requires that performance management become agile, decentralised, and political. Agile refers to the use of performance indicators, which should be used for learning purposes. As such, they can be adjusted to the context and the respective needs, and do not require the same stability as indicators for accountability do. Furthermore, in the ever-changing and fast-paced environment in which the public administrations now find themselves, perhaps the notion of “stable indicators” is somewhat outdated. Rather than building large, inflexible top-down performance management systems, Van Dooren (2011) argues for empowering middle managers and frontline supervisors who have a better knowledge and overview of the specific situations. Furthermore, in systems constructed in this manner, it would be faster and easier to provide performance information on demand – what Van Dooren (2011) calls the “guerrilla tactics” – provide information directly when needed. This, however, also implies that the users and providers of performance information are able to understand and apply this information. Finally yet importantly, the political aspect of performance management should be acknowledged. This implies engaging in a dialogue on the performance information with different stakeholders. It is hoped that this leads to more evidence-based decisions and shifts the focus towards outputs and outcomes. So much for double-loop learning in performance management systems.

Deutero, or triple-loop learning, then could even question the underlying strategy and values by attempting to find answers to the question of how we can learn to learn how to measure performance. Perhaps the more radical question is (why) do we need to measure performance? Triple-loop learning “assesses the ability to assess performance with the given performance indicators” (Marr, 2009, p. 215). It asks whether performance indicators are helpful for improving the organisation, and whether the organisation can learn from them.

As is now clear from the elaborations above, learning does not occur in a vacuum. Learning can take on different forms, which are likely to depend on the context in which learning takes place. This includes both structural and cultural aspects. Not every

organisation allows or actively enables learning. The next section focuses on the learning organisation as the “construction” that enables or facilitates learning, before the structural and cultural context is examined more concretely.

3.3.3 The learning organisation – an approximation to the concept

The learning organisation is

“skilled at creating, acquiring and transferring knowledge, and at modifying its behaviour to reflect new knowledge and insights. ... Learning organizations are skilled at five main activities: systematic problem solving, experimentation with new approaches, learning from their own experience and past history, learning from the experiences and best practices of others, and transferring knowledge quickly and efficiently throughout the organization” (Garvin, 1993).

“one that learns continuously and transforms itself. ... Learning is a continuous, strategically used process – integrated with, and running parallel to, work. Learning results in changes in knowledge, beliefs, and behaviours. Learning also enhances organisational capacity for innovation and growth. The learning organisation has embedded systems or mechanisms to capture and share learning” (Watkins & Golembiewski, 1995, p. 87).

“[an organisation] where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together” (Senge, 2006, p. 14).

“an entity, an ideal type of organisation, which has the capacity to learn effectively and thereby to prosper. ... [when analysing learning organisations, one wants] to understand how to create and improve this learning capacity, and therefore ... have a more practical impact and a performance agenda” (Easterby-Smith & Lyles, 2011b, p. 3).

These definitions point out that while organisational learning refers to the *process* of learning, the learning organisation describes the *form* of an organisation. At the same time, Watkins and Golembiewski's (1995) definition clarifies that a strict distinction between the two is not always possible (see also Easterby-Smith et al., 2000). Nevertheless, for the purpose of clarification, this section elaborates further on the concept of the learning organisation.

Probably one of the most prominent representatives of the learning organisation concept is the seminal work of Peter Senge on “The fifth discipline” (Senge, 2006). He develops a model of the learning organisation, consisting of five elements, the so-called five disciplines: Personal mastery, shared vision, mental models, dialogue, and systems thinking.

Senge describes *personal mastery* as the spiritual foundation of an organisation. It is “the discipline of continually clarifying and deepening our personal vision, of focusing our energies, of developing patience, and of seeing reality objectively” (Senge, 2006, p.

131). Furthermore, it is about “learning how to generate and sustain creative tension in our lives”. Creative tension is the contrast between “vision (what we want) and a clear picture of current reality (where we are relative to what we want)” (Senge, 2006, p. 132). Personal mastery lies at the heart of the learning organisation, as organisational learning is only possible when individuals learn. This is especially true for leaders and managers of an organisation. The “old dogma of planning, organising, and controlling” (Senge, 2006, p. 130) should be replaced by focusing on personal growth of all employees, inspiring them to create something bigger, to be creative in achieving their aims.

Thus, personal mastery is not, as perhaps mistakenly supposed, about “dominance” over people. It is rather a process of becoming more proficient in personal terms, a process of personal growth and awareness. Hence, it is primarily an individual achievement, visible in personal efforts of a member of an organisation.

However, personal mastery also has a critical potential. Empowering employees and making them think freely and creatively is not something to which every manager aspires. This might also have to do with the potential risks personal mastery entails, if implemented without the other disciplines of the learning organisation presented below. Especially shared vision and mental models are closely connected to personal mastery. Just imagine an organisation that consists of a considerable number of employees who all have developed “personal mastery”, but without the context of a shared vision or shared mental models. Rather than pushing the organisation forward to achieving its aim and purpose, this constellation might prove counterproductive in such an “unaligned organisation” (Senge, 2006, p. 136).

A *shared vision* is another discipline of the model, which refers to a true, genuine vision for the organisation or the team, to which all members want to commit, instead of simply complying with it. In an organisation with a shared vision, members are supposed to learn because they want to, as opposed to because they have to. A shared vision creates an overarching goal to which people are committed. By having that vision and the overarching goal in mind, people are encouraged to take risks and experiment in how to achieve that vision. There is no right or wrong; the important thing is that people are on their way, attempting what they can with the means available and the vision in mind. Importantly, Senge (2006) claims that a vision can only fulfil its potential if people genuinely believe that they can shape their future, or rather, the future of their organisation, as opposed to merely reacting to it. This is about “generating change, not reacting to change”.

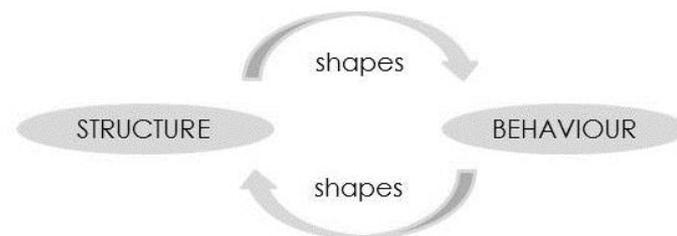
Mental models mean assumptions, generalisations, pictures, or images that influence how we understand the world and how we take action. Mental models are “deeply held internal images of how the world works” which eventually “limit us to familiar ways of thinking and acting” (Senge, 2006, p. 163). Thus, they shape our behaviour and our thoughts, which is not necessarily a bad thing. However, Senge (2006, p. 166) claims that it becomes problematic when these models become implicit, when we stop questioning them, and thus, are no longer aware of their existence. As a shift of mind lies at the heart of the learning organisation, again the move from the dogma of “managing,

organising, and controlling” towards “vision, values, and mental models” has to be restated (Senge, 2006, p. 171). Note here the link to double-loop and triple-loop learning that place challenging and questioning assumptions, beliefs, and values at their cores.

Dialogue is about team learning. When members of the organisation suspend their assumptions, meet each other as colleagues with the aim to think together, to build something significant, then genuine dialogue can happen. This is essential for building and sustaining learning organisations. A non-confrontational atmosphere in the organisation prevents “defensive routines” that impede organisational learning (Senge, 2006; see also Coopey and Burgoyne, 2000 and Oswick et al., 2000). This is also reflected in what has been established in the previous section, i.e. constructive, positive learning for change can only occur when favourable conditions within the organisation are established.

Finally, *systems thinking* is the overarching discipline for learning organisations (Senge, 2006). It is “a conceptual framework, a body of knowledge and tools”, and consists of two main elements that mutually reinforce themselves – structure and behaviour. Structure in this context refers to the set of beliefs, assumptions, established practices, skills, and networks of relationships that influence how people make decisions. Thus, structure shapes the behaviour of actors. At the same time, the behaviour of actors is based on underlying structures (see Figure 5). Hence, if we want to change the structures of the system, we have to commence with behaving differently.

Figure 5 The structure-behaviour relationship



Source: based on Senge (2006)

A more recent contribution to the research on the learning organisation is offered by Örtenblad (2013). He suggests a more systematic approach towards the learning organisation by specifying four different aspects/types of the definition of the learning organisation, which comprise the following: learning at work, organisational learning, climate for learning, and learning structure. Örtenblad categorises these definitions into the following two dimensions: Weak governance vs. strong governance, and processes vs. organisation form (see Figure 6).

Figure 6 Distinctions between the four aspects/types of the learning organisation

	Weak governance		
Processes	Learning at work	Climate for learning	Organisation form
	Organisational learning	Learning structure	
	Strong governance		

Source: Örtenblad (2013b, p. 24)

Weak governance implies definitions of the learning organisation which are more “humanistically oriented and include less control of people and what and how they learn”, while strong governance implies “managerially informed” definitions and more control (Örtenblad, 2013b, p. 24). Processes then refer to what I previously define as “organisational learning”, i.e. the processes of learning, whereas the organisation form corresponds to the foregoing discussion on the actual learning organisation. Interestingly, Örtenblad (2013) argues that organisational learning can be seen as part, or an aspect, of the learning organisation. He also refers to other scholars who contend the opposite, namely that the learning organisation is an aspect of organisational learning (e.g. Easterby-Smith et al., 2000). I propose that the concepts are interdependent. Organisational learning takes place in a learning organisation. However, a learning organisation refers to the learning *capacities* of an organisation. The simple fact that an organisation is described as a learning organisation does not guarantee that learning takes place. Organisational learning refers to the process of learning and, certainly, the learning organisation facilitates this learning to take place.

Learning at work refers to learning from more experienced others, learning by doing, and tutoring or mentoring. It is thus a process that takes place at work, not outside of it.

Organisational learning, according to Örtenblad’s definition, is the acquisition, storage, and use of knowledge, by means of single- and double-loop learning, and organisational memory (rules, routines, manuals, etc.).

A *climate for learning* then refers to the facilitation of individual learning, e.g. through creating space for experimentation and reflection. Learning is the result, and structure is the means to achieve that result. Thus, a climate for learning is one that allows for mistakes by considering them as learning opportunities rather than as failures.

The *learning structure* then considers learning as a means for achieving flexibility. This is likely in organisations characterised by flexibility, autonomy, decentralisation, empowerment, continuous learning, and a non-hierarchical structure. According to Örtenblad, Senge’s definition of and elaboration on the learning organisation falls under this aspect (Örtenblad, 2013b, pp. 25–30).

In summary, the learning organisation describes the form of an organisation, as opposed to the process. It is constituted by cultural and structural elements, such as a shared

vision, systems thinking, personal mastery, a focus on dialogue and interaction, a climate that allows for experimentation and failure. These structural and cultural elements shape the behaviour of the actors within the organisation. At the same time, the actors' behaviour shapes culture and structure. In the following section, some aspects of the organisation form, i.e. structure and climate, are examined in detail.

3.3.4 Context matters: Structure and culture

It is clear from the conceptualisations of the learning organisation and organisational learning, that learning is inseparably embedded in a wider organisational context (Easterby-Smith et al., 2000; Örtenblad, 2013; Richter, 1998).

Context factors that play an important role are, as mentioned, beliefs, norms, and values of an organisation (Mahler, 1997; Rashman et al., 2009; Senge, 2006). Furthermore, as Gilson et al. (2009), Örtenblad (2013b), Senge (2006), and other authors argue, human resources, leadership qualities, and team work abilities are key to a successful learning process. According to the social constructivist view, learners are understood as “social beings who construct their understanding and learn from social interaction within specific socio-cultural and material settings” (Easterby-Smith et al., 2000, p. 787). Learners gather together in dialogue and discussion, which fosters and produces organisational learning (Burgoyne & Jackson, 1997; Coopey & Burgoyne, 2000; Easterby-Smith et al., 2000; Oswick et al., 2000; Richter, 1998; Senge, 2006).

However, not only the cultural context plays an important role; the structural perspective claims that formal mechanisms, so-called organizational learning mechanisms (OLM), must be taken into account when attempting to determine how learning is achieved in an organization (Lipshitz, Popper & Oz, 1996). Organisational learning mechanisms are “the structural procedures by which individual learning is acquired and utilized for organizational purposes” (Moynihan & Landuyt, 2009, p. 1098). The structural perspective also includes Örtenblad's (2013b) types of learning organisation and climate for learning.

Moynihan and Landuyt (2009) contend that it is not a question of either/or, but rather one of both/and. Organisational learning consists of both structural and cultural aspects. These include information systems (the “organisational learning mechanisms”), adequacy of resources (adequate resources as facilitator of organisational learning), mission orientation (what Senge (2006) describes as shared vision and systems thinking), decision flexibility (linking learning to decisions by enabling teams to take part in decision-making), and learning forums. The latter are, according to Moynihan and Landuyt's (2009) analysis, the most influential factors and, as such, deserve special attention here.

Learning forums bridge cultural and structural perspectives on learning. Learning forums are “routines that encourage actors to closely examine information, consider its significance, and decide how it will affect future action” (Moynihan, 2005). Learning forums link the (performance) data collection and dissemination, and their use. Learning forums are where data is examined and interpreted. Thus, learning forums are a

moment of dialogue. Learning forums, furthermore, are established by formal rules, but they can reach their potential only if appropriate cultural features are present. In other words, both structure and culture must be considered when establishing and working with learning forums (Moynihan, 2005, 2011; Moynihan & Landuyt, 2009). An overview of the elements of learning forums by Moynihan (2005, p. 211) is provided below (see Box 2).

Box 2 Elements of learning forums

- Routine event
- Facilitation and ground rules to structure dialogue
- Non-confrontational approach to avoid defensive reactions
- Collegiality and equality among participants
- Diverse set of organisational actor responsible for producing the outcomes under review
- Dialogue centred, with dialogue focused on organisational goals
- Basic assumptions are identified, examined, and suspended (especially for double-loop learning)
- Quantitative knowledge that identifies successes and failures, including goals, targets, outcomes, and points of comparison
- Experiential knowledge of process and work conditions that explain successes, failures, and the possibility of innovation

Source: Moynihan (2005, p. 211)

Learning forums are the cornerstone of Moynihan's (2008) interactive dialogue model. This model is based on the following assumptions (Moynihan, 2008, p. 102):

- Performance information is not comprehensive.
- Performance information is ambiguous.
- Performance information is subjective.
- Production of performance information does not guarantee use.
- Institutional affiliation and individual beliefs will affect selection, perception, and presentation of performance information.
- The context of dialogue will affect the ability to use performance information to develop solutions.

An important cultural characteristic of learning forums is the emphasis on equal-basis dialogue. Interactive dialogue acts “as a social process that helps to create shared mental models, has a unifying effect, and helps to develop credible commitment for the execution phase” (Moynihan, 2011, p. 39). Here, we see the link to Senge's (2006) conceptualisation, and both authors emphasise the non-confrontational character of learning forums – all participants are equal and collegial. Learning forums work best when based on a purpose-driven dialogue, which allows for openness and different

opinions without creating situations of blaming. Senge (2006) stresses the distinction between discussion and dialogue, with the former being characterised as a “ping pong of ideas”, while the latter consists of a constructive conversation among equal participants. A non-confrontational atmosphere prevents “defensive routines” that impede organisational learning.

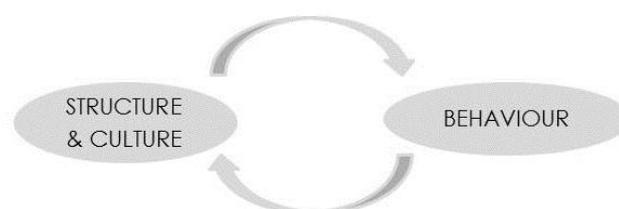
Similarly, Coopey and Burgoyne (2000) focus on the importance of “conflictual and creative argument” in critical decision-making, as “defensive avoidance” is a likely phenomenon likely in highly cohesive groups when decisions have to be taken under time pressure. To avoid this defensiveness, Coopey and Burgoyne stress the need for systematically questioning the reasoning processes (p. 878). Also Oswick et al. (2000, p. 888) claim that “discourse, as the dynamic and interactive process through which the dominant ‘organizational reality’ is socially mediated, is central to understanding organizational learning as a multi-layered intersubjective phenomenon.”

This line of thinking is also expressed in Burgoyne and Jackson's (1997) “arena thesis”. This thesis takes a pluralist perspective that views “management learning as a domain within which conflicting purposes and values within an organisation meet to be reinforced, reconciled or proliferated” (Burgoyne & Jackson, 1997, p. 61). Thus, the “arena” is a meeting place for different desires and interests, which are to be reconciled and reconfigured. Arenas are characterised as visible, accessible, and manageable.

To summarise, the scholarly literature (notably Senge, 2006, and Moynihan 2005, 2011) claims that a non-confrontational atmosphere in learning forums is key to success of the latter. A conflictual atmosphere, on the other hand, might impede organisational learning by evoking defensiveness. However, a non-confrontational atmosphere does not mean the absence of conflicts and diverging opinions. It rather refers to the behaviour among the forum participants. When learning forums are purpose-driven and open to constructively seeking alternatives in situations of conflict and disagreement, they are more likely to be successful (e.g. Coopey & Burgoyne, 2000). The main common element in these views on learning forums is the emphasis on constructive dialogue (Burgoyne & Jackson, 1997; Coopey & Burgoyne, 2000; Easterby-Smith et al., 2000; Oswick et al., 2000; Richter, 1998; Senge, 2006).

This section has shown that for genuine learning to occur (and to be enabled), both structural and cultural factors have to be taken into account. Figure 5 can thus be extended by the cultural dimension (see Figure 7):

Figure 7 The structure-culture-behaviour relationship



3.4. Accountability and learning in performance management systems?

The previous chapters show that, on the one hand, we have ample evidence that indicator-based accountability for performance does not live up to its expectations, and, more severely, even contradicts or reverses it at times. On the other hand, while learning systems appear to be a forward-looking alternative, despite its criticism, it would be unrealistic to assume that accountability purposes can simply be replaced by learning purposes, for accountability in the sense of responsibility or answerability is indispensable in any democratic society. Furthermore, section 3.2.2 elaborated on the “paradox” concerning accountability, resulting in a likely incompatibility between accountability and learning purposes. So which way to take in this dilemma?

Recently, several authors have proposed suggestions. Van Dooren and Willems (forthcoming) recommend shifting accountability towards innovation rather than towards results. Analogously, performance measurement would then also shift towards learning. As such, we would have “loosely coupled systems” of performance measurement and accountability, as outlined above (see section 2.5). Van Dooren and Willems (forthcoming) claim that “performance information will only lead to learning and innovation when the pressure emanating from indicator-based accountability is driven under the gaming threshold”. Thus, performance information must be decoupled from formal accountability processes such as budgeting or contracting. The challenging task would be to find performance indicators that trigger learning. Furthermore, the authors emphasise the need for dialogue of (meaningful) results.

By the same token, Mayne (2007, p. 66) contends, in line with other scholars, that accountability should “move away from a process of shaming towards one of learning and improving from past experiences”. Nevertheless, he also admits that this is probably one of the most significant challenges.

Similarly, Olejniczak and Newcomer (2014) advocate a shift towards accountability for learning. They elaborate on the necessity to develop “an error-embracing culture in the public sector” (Olejniczak & Newcomer, 2014, p. 91). Especially in such a complex setting as the public sector, they argue, it is indispensable to redefine failure and create space for experimentation. They suggest classifying “failures” on a spectrum of blame-worthy to praise-worthy, and acting on them accordingly. In this way, exploration, for example, in the form of pilot studies, could be encouraged, as it would be seen as a “blame-free space for experimentation”, while severe “failures”, such as conscious violation of procedures, could be addressed immediately.

Furthermore, Olejniczak and Newcomer (2014) suggest to refocus accountability to promote learning. This implies that “public leaders and managers be held accountable for understanding, learning, and for implementing lessons learned” (p. 93). As such, the consequences, rather than the processes, of public interventions would be judged. This would shift the focus to the bigger picture. However, this also implies that not every consequence can be measured immediately. For certain effects to materialise, it might

take time. Thus, a flexible approach is warranted. Next, Olejniczak and Newcomer (2014) suggest public managers and leaders be held accountable for learning in three main ways. They should be accountable for

- actively looking for solutions for observed processes,
- intentionally exploring the context during implementation, and
- implementing lessons learned.

This approach challenges the traditional notion of performance accountability substantially in that it shifts focus towards “constructive, incremental building of understanding of patterns, structural processes and underlying mechanisms” as opposed to ad-hoc reactions (Olejniczak & Newcomer, 2014, p. 94).

Guijt (2010) elaborates on the “myth” of the incompatibility of accountability and learning: “You cannot be accountable if you do not learn. And you need to know how well you live up to performance expectations in order to learn.” (p. 21). Importantly, both accountability and learning are endeavours that aim at dialogue and making sense of (performance) information. However, she admits that, while this sounds convincing in theory, practice has not lived up to this mantra, and that in particular rigid accountability systems inhibit learning. Thus, we have to move towards “learning-oriented accountability – or accountability that feeds into learning” (Guijt, 2010, p. 289).

Establishing a direct link between performance measurement and accountability and learning, Yang and Holzer (2006) expand on performance measurement as a tool for both learning and accountability. Performance measurement has the potential to facilitate double-loop learning, when reformed adequately. Performance measurement “must be reformed to construct a real accountability and learning instrument that goes beyond the input – output model, an instrument that is more appropriate for mechanical managerial control.” To achieve this, both evaluators and the evaluated have to be included in the process, which essentially relies on participation, interaction, and communication (Yang & Holzer, 2006, p. 123).

The latter is also put forward by Kettl (2001). He finds that “performance is much more about communication than measurement. [And] performance measurement is too important to be considered as simply a measurement process or as a governance process that can be left solely to administrators” (Kettl, 2001, pp. 8–9). Thus, he argues for a focus on communication, not measurement; the quality and usefulness of performance measures lie in their use for management and policy decisions. Thus, performance measures must lead to performance management. In turn, performance management is more than performance measurement. The dialogue and discourse following from the performance measures is what makes performance measurement a useful, constructive endeavour.

In addition, Greiling and Halachmi (2013) argue that “the adequacy of efforts to ensure accountability should be measured or evaluated by reference to their actual use to generate organizational learning, and not by their sheer numbers or sophistication.” Interestingly, the authors make the point that accountability can ensure that “things are

done right”, but it cannot ensure that “the right things are being done”. The latter, however, is the one that counts in the long-term. (Recall the distinction between single- and double-loop learning.) Following this line, the authors contend that while learning is focused on the long-term perspective, accountability is a short-term evaluation. To bridge this gap, Greiling and Halachmi frame the concept of “dynamic accountability”, which “results from organizational learning that derives lessons that improve the definition of goals, organizational structure, standard operating procedures, and subsequent collection of data” (Greiling & Halachmi, 2013, p. 384). Furthermore, they make the distinction between formal and informal accountability and contend that both have their disadvantages. Formal accountability can lead to data manipulation, gaming, and cheating; informal accountability can lead to problems of “accessibility, overload, biased predispositions, or individual idiosyncrasies” (Greiling & Halachmi, 2013, p. 401). However, they argue that these considerations are mainly theoretical – empirical evidence for linking accountability and learning is lacking.

Schillemans et al. (2013) examine traditional and “innovative” forms of accountability and their potential for learning efforts. They conclude that traditional accountability differs substantially from innovative accountability (see Table 4). While the authors generally do not exclude learning from traditional accountability, they are more confident that innovative accountability can enable or foster learning. For instance, innovative accountability emphasises different sources of information, depending on the respective needs and context. Most important for the learning debate is the fact that innovative accountability focuses on the debating phase of the accountability process, as opposed to the focus on the information provision phase of traditional accountability efforts. Schillemans et al. (2013) argue that the crucial element here is decoupling accountability mechanisms from strict sanctions. This would enable reflective dialogue and debate. Similarly, also Brignall and Modell (2000), Johnsen (1999), Laegreid, Roness and Rubecksen (2008), and Van Dooren et al. (2015) plea for decoupling indicators from goals (and as such, from sanctions in case of not reaching the goals). Importantly, however, while these are – in theory – notable efforts to bridge accountability and learning, reality shows that despite all these elements of the ideal type of innovative accountability being present, it is still not a given that learning will in effect occur. For example, debates could be intimidating, blame could be shifted towards lower hierarchical levels, and even within the debate forums, deconstructive “grilling” could take place (Schillemans et al., 2013, p. 426). Here we can see a need for Moynihan’s “learning forums” and Senge’s idea of the learning organisation, where exactly such behaviour is prevented by creating forums of debate in which participants are on equal footing and work together to a shared vision.

Table 4 Traditional and innovative accountability compared

Traditional accountability	Innovative accountability
Standard, permanent structures	Made-to-measure, temporary structures
Wide, generic, comprehensive	Deep, specific, selective
Cyclical	Incidental
Financial focus	Focus on policy outcomes
Vertical orientation	Horizontal orientation
Focus on information provision	Focus on debating phase of accountability
Sanctions phase: judgement and correction	Sanctions phase: learning and improvement

Source: Schillemans, Twist & Vanhommerig (2013, p. 425)

While these contributions reflect mostly theoretically on the question of whether accountability and learning can go together, Schillemans and Smulders (2016) look more deeply into the questions of whether public managers actually learn, what, and under which conditions. While the authors do not elaborate extensively on the “whether” question, they find three conditions that are – theoretically – conducive for learning: 1) actors being held accountable in a forum that has a long-term focus rather than a strict control and regularity focus; 2) interpersonal trust among account-holders and account-givers; and 3) the availability of formal sanctions. Regarding the latter, the authors concede, however, that sanctions can be supportive as well as destructive for learning. It is thus a question of finding the right “middle way”. Nevertheless, as has been showed in chapter 2 on the effects of the use of performance information, the threshold for gaming and cheating is low, when sanctions are possible and real. Regarding the question of what actors can learn, Schillemans and Smulders (2016) suggest three main objects: knowledge, skills, and ideas. Knowledge and/or understanding of internal and external developments; managerial skills, both general and specific with regard to the accountability processes; and ideas about strategic and operational issues. Based on their empirical analysis in the Dutch context, the authors conclude that both control and critical feedback are necessary for learning from accountability to occur. However, they also admit that this is especially true for an egalitarian, not-politicised context, and thus, for administrative, structural forms of accountability. A different picture is likely to emerge when we examine politicised and hierarchical contexts, where the level of trust among account-holders and account-givers might be lower, and the results of policies or programmes are under much higher scrutiny.

Molleman and Timmerman (2003) argue that if performance management was to focus more on knowledge creation, learning, and innovation, this would have major implications, in particular for human resource management. Focus would have to shift to the personal and interpersonal levels. In fact, team processes would become much more important than they probably are nowadays. The authors find that to bring performance management in line with a genuine learning and innovation policy, several changes have

to be considered. For instance, focus more on the interpersonal, rather than the individual level. Furthermore, either reward separately from the performance management system, or link rewards with the outcomes of a project team as a whole, rather than with individual successes. Dedicate attention also to qualitative aspects of performance. Include more intrinsic and less extrinsic motivators in the reward system. Although the authors test their theoretical model in a private sector context (R&D departments of a firm), based on the previous literature research, I see a clear benefit for the discussion at stake, as this contribution also explicitly refers to the implications for human resource management, which also plays an important role in the public sector.

In conclusion, this section has shown that there is a lively debate about the compatibility of learning and accountability in performance management systems. Some authors (Guijt, 2010; Yang & Holzer, 2006) are in theory of the opinion that one could feed the other. However, they admit that empirical evidence for this outcome is lacking. Hence, an increasing number of authors suggest middle ways for accountability and learning to co-exist in the same system (Greiling & Halachmi, 2013, 2013; Mayne, 2007; Olejniczak & Newcomer, 2014; Schillemans et al., 2013). These middle ways usually imply a form of “accountability for learning”. In essence, this means that performance information would be decoupled from sanctions, and as such, used in a “soft way”. Room should be provided for experimentation, and failure should be considered as a form of learning to improve. Consequently, account-givers should be held to account for their efforts to induce learning, experimentation, and innovation. Key to any such endeavour would be communication and dialogue among the actors involved. Importantly, these considerations imply a “radical” re-thinking not only of “accountability” which would not be based on indicators anymore, but also of the way performance has been managed to date.

3.5. Accountability, learning, and trust

The previous section has found that hierarchies are likely to inhibit learning, while trust is considered to be conducive for learning. This section thus has a closer look at the concepts of trust and power relations in the context of accountability and learning.

Trust can be defined as “the intentional and behavioural suspension of vulnerability by a trustor on the basis of positive expectations of a trustee.” (Oomsels & Bouckaert, 2014). Bouckaert (2012) categorises trust along three dimensions: Trust of citizens and organisations in government and the public sector (T1); trust of government and the public sector in citizens and organisations (T2); and trust within the public administration, which includes civil servants, public sector organisations, and politicians (T3). When we examine the relationship between trust and accountability and learning, we are thus primarily concerned with T3, trust within the public administration. However, as trust is not the main focus of this thesis, an elaborate discussion on trust and its dimensions would divert attention from the actual topic under investigation. Nevertheless, this information should suffice as background information when looking at scholarly contributions to the link between trust, accountability, and learning.

According to O'Neill (2002), in relation to the discussion about accountability and trust, it is rather trustworthiness we have to consider. Accountability systems have been set up with the aim to provide evidence of trustworthiness of politicians, civil servants, and institutions. Moreover, to show trustworthiness, politicians, civil servants, and institutions have to show their vulnerability.

However, O'Neill (2002, 19) asks critically whether our accountability efforts have not actually deteriorated trust:

“Perhaps the culture of accountability that we are relentlessly building for ourselves actually damages trust rather than supporting it. Plants don't flourish when we pull them up too often to check how their roots are growing: political, institutional and professional life too may not flourish if we constantly up-root it to demonstrate that everything is transparent and trustworthy.”

A similar claim is put forward by Johansson and Montin (2014). They address the complex relationship between accountability and trust, particularly in high trust societies. While the underlying premise of NPM inspired performance management systems is that accountability mechanisms promote and stabilise trust, the authors argue that empirical evidence has shown that performance accountability might in fact erode trust. Factors that contribute to eroding, rather than reinforcing trust, are, for example, decision-making, which is based on rules rather than knowledge, a focus on deviation, with a naming and shaming policy resulting from it, and an abundance of indicators, whose quality is rarely challenged. Also the fact that politicians are becoming more reactive instead of proactive, due to performance management based on the market logic, disconnects citizens from politicians, and as such, increases the likelihood for distrust (Johansson & Montin, 2014).

This claim links to Möllering's (2005) elaboration on the link between trust and control. Rather than considering these concepts as dualism (either – or), Möllering suggests to regard them as duality. This perspective implies that “trust and control each assume the existence of the other, refer to each other and create each other, but remain irreducible to each other.” (Möllering, 2005, p. 284). This has implications for any analysis of trust and/or control relationships: One cannot be understood without the other; and the socially constructed structures within which these relationships play out, also play a non-negligible role.

Furthermore, the discussion on the relationship between trust and control reflects similar considerations as between accountability and learning, and can thus serve as illustration of the dilemma between the two concepts. Recalling the earlier elaboration of accountability and control, Costa and Bijlsma-Frankema (2007, p. 396) define “formal control” as an “approach [that] focuses on the establishment and utilization of formal rules, procedures, and policies to monitor and reward desirable performance”. Informal control, on the other hand “emphasises the regulatory power of organizational norms, values, culture, and the internalization of goals to encourage desirable outcomes.” These mechanisms are also present in the definition of accountability presented in section 3.2.1. Several authors examine the relationship between trust and control. The main

distinction is made between considering these two concepts as substitutional or as complementary (Bijlsma-Frankema, 2005; Costa & Bijlsma-Frankema, 2007; Oomsels, Porumbescu, Bouckaert & Im, 2013). Möllering (2005) bridges this “duality” by proposing a “middle way” in regarding trust and control as dualism.

However, trust is not only important when we examine accountability mechanisms in organisations. Trust plays an important role also in the context of learning mechanisms.

Rashman et al. (2009) elaborate on several context factors that influence the extent to which learning can take place. Mutual trust is one of them. However, they remark that while a certain level of trust is necessary to establish a blame-free space and supportive climate for cooperation, too much trust can lead to following the organisation’s path without questioning or reflecting critically on shared views. The same applies to power and influence; on the one hand, power and influence can help to create stability and gain access to resources and expertise, and as such, foster organisational learning. On the other hand, power and influence can also inhibit learning when actors withhold or manipulate their knowledge. Hence, trust and learning and power and learning relationships cannot be described in black and white terms. Nevertheless, it is widely held that trust is a positive contributor to organisational learning, as well as cross-boundary networking and risk-taking. As for power, research suggests that decentralisation and low or informal hierarchies support a learning environment, whereas hierarchical power structures have the opposite effect. A strong blame culture, for example, is likely to have a negative effect on learning (Rashman et al., 2009).

Also Bouckaert and Peters (2002) consider trust to be an important factor in bridging the gap between accountability and reporting on the one hand, and performance (improvement) and learning efforts on the other hand. This, however, asks for careful consideration as well. Too much trust and too much control can both have negative effects. As such, Bouckaert and Peters (2002) conclude that what is needed is “an appropriate mix between trust on one hand and performance-based guidance, control, and evaluation on the other hand.”

However, this idea of “letting and making managers manage”, essential to the NPM doctrine, has been considered problematic. As has been shown in the literature reviewed for this chapter so far, evidence for a co-existence of such improvement-based and control-based elements in performance management systems is scarce. Even more, in particular the presence of strong control elements in performance management systems does not leave enough room for trust- or improvement-based elements to play out (e.g. Norman, 2001).

Hence, the literature on the role of trust in accountability- and learning systems uniformly stresses the importance of the extent of trust. Too much trust can be as detrimental as too little trust. Nevertheless, generally, accountability systems are considered able to damage trust, as focus shifts towards following the rules, a naming and shaming policy, and a more reactive instead of proactive behaviour of politicians. Learning systems on the other hand are based on trust, which is created through cooperation and providing a blame-free space. Furthermore, it is found that hierarchies

might have a negative effect on trust. However, the research analysed also shows that it is not a story of either-or; trust is important in both accountability and learning systems. As such, it is a question of the extent of trust, which increases the likelihood for a stronger focus on accountability or learning. For this reason, in high-trust environments we can expect more elements of learning systems, while low-trust environments are likely to be based more on accountability elements.

3.6. Implications of the use of performance information for the design of performance measurement systems

Now that we have clarified the different ways performance information can be used in, another important aspect we have to examine is what the implications of the different forms of use are for the design of a performance measurement system. In an “ideal scenario” the use of performance information determines the design of the system. Or, framed differently, form follows function (Van Dooren et al., 2015).

We can think of the design of a performance measurement system several ways. Bouckaert and Halligan (2008) provide an overview of the changes of design over time, which they describe as “ideal types” of performance management.

In the beginning of performance management, performance measurement systems were created *ad hoc* and internally. The criteria for performance indicators were that they be “accurate, definitive and reliable” (Bouckaert & Halligan, 2008, p. 86). Hence, performance indicators were used in an objective and technical way.

In the following stage, it was recognised that measurement affects the behaviour of individuals or organisations. It is here, where dysfunctional effects – as outlined in chapter 2 of this thesis – start to become apparent. To counter these dysfunctional effects, efforts are undertaken to make the system more interactive by involving more stakeholders in its design. In addition, the quality and functionality of indicators, not only the technical requirements, became an important issue. This implied measuring not only inputs, but also outputs and outcome; not only tangible services and outputs, but also intangible ones. Furthermore, performance measures began to be used for punishments and rewards, and as such, lost their previous “observatory” role. Rather, performance indicators were used intentionally and purposefully. Another feature of this phase is that performance measurement systems were more systematically designed and implemented. This included the development of systematic methodologies to construct these systems, typologies of outputs, and the application of frameworks such as Total Quality Management (TQM) or other international standards, such as the ISO standards. In this context, also the purpose of “improvement” began to play a role. Several authors suggested step-models for the design of performance measurement systems (Behn, 2003; Pollitt, 1995; Van Dooren et al., 2015). While in this stage there is more potential for double-loop learning than in the previous stage, which was based on single-loop learning, Bouckaert and Halligan (2008, p. 99) find that this potential is impaired due to a “fragmented and disconnected focus of management”.

In the following ideal type of performance management, performance information “is systematically and coherently generated, integrated and used” (Bouckaert & Halligan, 2008, p. 100). The authors claim that this type is equally trust-based and performance-based, “even if there are tensions”. Legitimacy of performance indicators is added to technicality and functionality as requirements for indicators. This implies that the importance of the context in which the measurement system is placed as well as political interests influence the choice of performance measures. Performance information is used more in a political way, and as such, is not “neutral” anymore. Hence, the use of performance information is closely linked to the design of the performance measurement system. Ideally, in this type of system performance information is used to systematically compare results, to learn to improve, and to change, internally and externally. Thus, also here we could expect single- and double-, and to a certain extent, triple-loop learning. Furthermore, legitimacy also requires that more stakeholders are involved in the design and the implementation of the system. In addition, independent audit institutions become important in this context. In other words, the system has to be more open, both internally and externally. This can be achieved by combining top-down and bottom-up approaches (Bouckaert & Halligan, 2008).

As a last ideal type, Bouckaert and Halligan (2008) elaborate on performance governance. Performance governance emphasises the effect on society, and citizen engagement, a whole of government approach, and societal indicators become important. Here, a combination of single-, double-, and triple-loop learning is envisaged.

This distinction of different ideal types makes clear that performance measurement and management are no static concepts, but are constantly evolving. From top-down, ad-hoc designed systems to whole of government approaches with the participation of a variety of stakeholders, from input indicators to indicators which (try to) measure the impact on society, from single-learning to triple-learning efforts – with every new type new opportunities, but also new challenges emerge. While it was hoped that trust-based and control-based elements could feed each other in the same system, chapter 2 of this thesis has found that reality generally does not meet this ideal type of performance measurement.

Nevertheless, this development shows that the purpose governments pursue with performance measurement has a substantial influence on the design of the performance measurement system. It is thus only logical that the design of a performance measurement system with mainly a learning purpose looks different from the design of a system which focusses mainly on accountability (Van Dooren et al., 2015).

A first, important distinction is mentioned earlier, namely the distinction between hard and soft use of performance information. When performance measurement is tightly coupled to judgement, we speak of hard use. This is particularly the case in accountability-based systems; the account-givers are judged based on performance results. In learning systems, on the other hand, performance measurement is loosely coupled to judgement. Performance information is analysed and interpreted before it is judged upon. Importantly, the judgment is not *per se* indicative for soft or hard use of

performance information. Rather, we have to examine the impact of the judgment to be able to characterise use as hard or soft (Van Dooren et al., 2015).

When we now look more closely at the different design parameters for learning and accountability systems, the following differences emerge: In learning systems, measurement is undertaken to determine problems and consequently, solutions. Accountability-based systems are primarily concerned with societal visibility or complying with predetermined indicators. This brings us to the next step, the selection of the indicators. In learning systems, we find all kinds of indicators, from input to outcome indicators. In accountability systems, on the other hand, focus is laid on outcome or output indicators. Furthermore, the data collection differs. While, in learning systems, data from external observers is considered less useful, as learning is primarily an intrinsic motivation, in accountability systems this data can be useful. On the other hand, while surveys and self-assessments can be insightful when measuring outcomes of an organisation and internal processes of learning systems, accountability systems can generally not make much use of such data sources. Regarding the data analysis, learning systems can refer to a variety of analysis techniques, such as comparisons with own (past) performance, scientific standards, or the performance of other organisations. Important, however, is that the assessment is actionable. This means that, for example, aggregated indices are less helpful. The latter are, in turn, often helpful for accountability systems, as they provide performance information “at a glance”. Finally yet importantly, also the reporting format ideally differs in learning and accountability systems, as the target group is different. As outlined earlier, learning systems are more internally focussed, while accountability systems have a stronger external focus. Consequently, the main target group will be the own staff. However, also executive politicians or the public might be interested in certain information. Hence, in learning systems the reporting format would be oral witnesses or specialised reports. The target group of accountability systems, on the other hand, will be mass media, and consequently, journalists. In addition, accountability to supranational institutions plays an important role. Hence, annual reports, the budget, and accounts are typical reporting formats (Van Dooren et al., 2015). Of course, the top management is an important instance to report to – in both learning and accountability systems. The previous sections show that the choice for accountability or learning-based systems also has an impact on the top management’s role. Does the top management hold the organisation or unit to account for their results or for their learning efforts? The reporting format highly depends on the management style.

Table 5 below summarises the main distinctions between the design of learning and accountability-based systems.

Table 5 The design of the performance measurement system according to its use

	Learning	Accountability
System development & implementation	Top-down and bottom-up	Top-down
Ownership of the system	Variety of stakeholders, all levels of management	Primarily top management
Performance indicators	Flexible, adjustable, context-dependent Actionable Input, output, outcome	Stable & robust Reliable Output, outcome
Data collection	Surveys, self-assessments, Own, internal data External "objective" data (statistical data, research institutions)	Surveys Own, internal data External observers
Data analysis	E.g. comparisons with own (past) performance, scientific standards, or the performance of other organisations	E.g. comparisons with other organisations, scientific standards, Aggregated indices
Reporting format	Specialised reports	Annual reports (budget, accounts, etc.)
Target group	Own staff; executive politicians, public Top management	Mass media/journalists, supranational institutions Top management

3.7. Conclusion and implications

This and chapter 2 have provided the theoretical and conceptual backbone against which the empirical case study is set up. It is thus worth pausing for a moment and summing up the main theoretical insights so far.

First, performance management systems have originally been built to increase efficiency and effectiveness of public administrations. "Doing more with less" – this is an often-cited underlying idea of performance management systems. This includes linking performance indicators to results, and consequently to either sanctions or rewards. Actors within the system are held to account for their processes and results by their superiors or line managers.

However, accountability also takes place outside the administrative forum, such as in parliamentary debates, in the public sphere (media) and in certain circumstances in the market or legal forums. This has spurred creative ways of dealing with performance

indicators and results to avoid blame, which has led to severe (and less severe) cases of data manipulation, gaming, and cheating. In consequence, accountability of performance management systems, in particular indicator-based accountability, appears to have reached its limits. At the same time, accountability is a necessary cornerstone for democratic governance. So what to do?

A recurring suggestion in the literature is to change ways towards learning systems. Learning systems are based on dialogue and horizontal cooperation, as opposed to hierarchical, top-down command-and-control systems, as accountability systems are generally characterised. In particular, as we find ourselves in an ever more complex and fast-paced environment, (quantitative) performance indicators are unlikely to capture this complexity. Performance information is ambiguous, subjective, not comprehensive, and affected by institutional affiliation and individual beliefs (Moynihan, 2008, p. 102). It is indispensable to acknowledge these features of performance information instead of suppressing them. Performance information has to be made sense of. Hence, it is essential to analyse it, give room for interpretation, and consider its specific context to which the information is bound. This, however, is generally not given in strict accountability systems where performance information is tightly coupled to judgement.

As a middle way, the contributions to this topic suggest that performance indicators be decoupled from accountability. Performance indicators can serve as one (among many) information sources which are dealt with in, e.g., learning forums. Accountability, in turn, would not be based on indicators, but would focus on learning efforts.

This, however, has also important implications for the design of performance measurement systems. The underlying rationale of learning- and accountability systems is different. The difference in purpose and use of performance information should also be reflected in the design of the performance measurement systems. This chapter has provided a brief overview of main differences and characteristics of accountability- and learning systems. Important elements are, e.g. how and by whom the system is developed and implemented (top-down vs. bottom-up), who the “owner” of the system is, what kind of performance indicators are used, how data are collected, analysed, and reported, and what the target group is.

Learning can occur in the form of single-loop (learning how to do the same things better) or double-loop (learning how to do it differently) learning – or a mix of both types. Triple-loop learning describes the process of “learning to learn”. Learning is furthermore embedded in the structural and cultural context of an organisation. For learning to occur, a combination of these structural and cultural aspects has to be present in an organisation. Cultural aspects that are favourable for learning are, for example, high employee empowerment, participation, and discretion. In addition, a high-trust environment is favourable for learning efforts. A punishment-oriented control system and a strong top-down culture on the other hand are likely to inhibit learning. Moynihan (2008) considers learning forums as an adequate stage where cultural and structural aspects come together. Learning forums put dialogue among forum participants centre stage. This dialogue is likely to occur when a non-confrontational approach is followed,

and participants consider themselves as equal. Regarding the structural aspects, learning forums are characterised as routine events, during which both quantitative knowledge and experiential knowledge are considered and discussed.

Moreover, it is this soft use of performance information, i.e. the decoupling of performance information from judgement and consequences, and mediating it by dialogue, which is diametrically opposed to the underlying assumptions of indicator-based accountability systems. The latter are based on (direct) consequences and sanctions resulting from performance results. Furthermore, as we have seen, the threshold to engage in gaming and cheating activities is low. This is particularly the case when incentives to perform well are high and actors try to avoid blame. Various authors thus question that using performance information for accountability purposes is compatible with using it for learning purposes.

However, while theoretical contributions to the move from accountability to learning systems and this (in)compatibility of accountability and learning purposes are abundant, empirical evidence is comparably scarce. The empirical part of this thesis (chapter 6) adds to filling this gap. The next chapter outlines the methodological framework of the empirical study.

4. METHODOLOGICAL APPROACH

'Would you tell me, please, which way I ought to go from here?' Alice speaks to the Cheshire Cat. 'That depends a good deal on where you want to get to,' said the Cat.
~ Lewis Carroll, *Alice in Wonderland*

The methodological choices for the empirical analysis centre on the following three questions: what is analysed, why it is analysed, and how this is analysed? This chapter provides answers to these questions in the next subsections. The last section reflects on the case study method in general, and the particular methodological choices taken for the present research.

4.1. A “most likely/least likely” comparative case study – the what and the why

To analyse the use of performance information in performance management systems, I have opted for conducting a comparative case study with a theory testing purpose. Here, I follow Flyvbjerg (2006, p. 227) in conceiving ‘theory’ in a soft sense, i.e. “testing propositions or hypotheses”, as opposed to “comprising explanation and prediction”. Due to the complex and multi-dimensional issue at stake, I believe that an in-depth multiple case study, which offers the possibility to include various dimensions and points of view, is the most adequate form for addressing the matter at hand. Furthermore, I see this case study as a means to an end – to illustrate the complexities of the use of performance information. As such, this case study is of instrumental, rather than intrinsic value (Stake, 1995; Thomas, 2011). By analysing multiple cases, in addition, I have the opportunity to explore and understand similarities and differences across cases.

More concretely, I opt for a most likely/least likely case design as this design is particularly useful for testing theoretical propositions, prove or falsify these propositions, or demonstrate their limits (Flyvbjerg, 2006; Gerring, 2007; Levy, 2008; Saetren, 2012). Most likely/least likely cases follow the often-cited “Sinatra” inference: Least likely cases assume that “if I can make it there, I can make it anywhere”. Most likely cases, on the other hand, depart from the position that “if I cannot make it here, I cannot make it anywhere” (Levy, 2008, p. 12). Thus, most likely cases are those, which, based on theoretical or analytical models, we expect most likely to confirm our assumptions. Contrary, least likely cases are those that are least likely to confirm our assumptions, and as such, most likely to disconfirm them (Flyvbjerg, 2006; George & Bennett, 2005; Levy, 2008).

From the literature reviewed above, certain drivers can be crystallised that, when present, presumably foster learning processes in an organisation. For instance, when an organisation consists of a culture with low hierarchies, a high-trust environment, organises regular meetings or forums where participants are treated equally and exchange experiences and knowledge, we can expect this organisation to envisage and apply learning, as opposed to accountability processes. Contrarily, when hierarchical, legalist structures and a low-trust environment prevail, we expect accountability to be the focus of performance management efforts.

An example for a most likely case for performance management systems for learning is the Finnish performance management system. The Finnish administrative culture is characterised as egalitarian, with a comparably low power distance according to Hofstede's model of cultures (Hofstede, 2001). It is also defined as collectivist, as opposed to individualist, and members of the culture are integrated into strong groups. Furthermore, Finland's culture is typically oriented towards long-term goals. Thus, according to these characteristics of the administrative culture in Finland, we can expect learning, rather than accountability elements, in the first instance to be present in the performance management system design. Hence, I conceive Finland as a "most likely" case; it is very likely that we find learning elements in the Finnish system. If, however, we cannot find them, it is probably difficult to find them anywhere else (Flyvbjerg, 2006, p. 229).

The Lithuanian performance management system, on the other hand, is an example for a least likely case. It is characterised by a hierarchical and legalistic culture (Nakrošis, 2001). Hence, we can expect the system to be based primarily on accountability structures, and as such as a least likely case for learning efforts (while being a most-likely case for accountability purposes).

Both cases are described in more detail in chapter 5.

The empirical part of this thesis is thus based on the following **propositions**.

Accountability and learning purposes are unlikely to be compatible in performance management systems.

- 1) Finland, based on its administrative structure and culture, is a most likely case for learning elements in its performance management system.
- 2) Lithuania, based on its administrative structure and culture, is a least likely case for learning elements in its performance management system.

To test these propositions, the following **research questions** are addressed:

How do performance management systems deal with the tension between accountability and learning purposes of the use of performance information?

- 1) What are the implications for performance management systems of the extent to which accountability elements are used?
- 2) What are the implications for performance management systems of the extent to which learning elements are used?
- 3) Are there indications of incompatibilities between the two?

4.2. Analytical approach: Qualitative interviews – the how

Kvale (1996) described the idea behind methodology as a “route that leads to the goal”. To keep up with this travel notion, I quote Kvale's (1996) travel metaphor in relation to conducting interviews:

“[The travel metaphor] understands the interviewer as a traveller on a journey that leads to a tale to be told upon returning home. The interviewer-traveller wanders through the landscape and enters into conversations with the people encountered. The traveller explores the many domains of the country, as unknown territory or with maps, roaming freely around the territory. The traveller may also deliberately seek specific sites or topics by following a method, with the original Greek meaning of ‘a route that leads to the goal.’ The interviewer wanders along with the local inhabitants, asks questions that lead the subjects to tell their own stories of their lived world, and converses with them in the original Latin meaning of conversation as ‘wandering together with.’” (Kvale, 1996, p. 4).

Moreover, quite literally, I have been such an interviewer-traveller, by visiting both Finland and Lithuania and conducting interviews with the respective stakeholders there. Now, as I have “returned home”, I am about to tell their stories.

This analysis emanates from qualitative interviews conducted with current and former civil servants and consultants of Finland and Lithuania (see table 6 below). This approach has been chosen, as it offers to understand and integrate multiple points of view, to develop detailed and holistic descriptions, and to gain an insider's perspective (Luton, 2010, p. 25; Weiss, 1995).

Table 6 Overview of organisations and respondents interviewed

FINLAND	Respondent	LITHUANIA	Respondent
Prime Minister's Office	Programme Counsellor	Office of the Government Sunset Commission	Advisor Head of division, Advisor
Ministry of Finance	Legislative Counsellor	Ministry of Finance	Deputy Department Director
Ministry of Transport and Communication	Controller	Ministry of Labour and Social Security	Department Director, Advisor
Ministry of Justice	Unit Director, Inspector General	Ministry of Education and Science	Head of unit
HAUS (Finnish Institute of Public Management)	Training Managers		
Audit Committee of the Finnish Parliament	Committee Counsellor		
National Audit Office	Deputy Auditor General, Performance Audit Counsellor	National Audit Office	Deputy Director
Controller, Ministry of Finance	Deputy Financial Controller	Ministry of Economy – Business inspection agency	Expert of Business optimization project
BroadScope Management Consulting	Senior Partner	ESTEP (private consultancy)	Member of the Board
Finnish Customs / Ministry of Finance	Director, Financial Manager	Lithuanian Customs	Chief inspector, Head of division
State Treasury	Service Manager	Ministry of the Interior	Department Director, Advisors

Interviewees were selected both purposefully and according to the snowball principle (Luton, 2010, p. 39); purposefully in the sense that in both Finland and Lithuania the interviewees are experts in their respective fields concerning performance management, and the intention was to include a “maximum range of perspectives” (Luton, 2010, p. 39). As I had three months “in situ” to conduct my interviews for the Finnish case study, I started with interviewing officials I was recommended to talk to by (Finnish) academics, and from there I followed according to the snowball principle, i.e. getting further recommendations either from interviewees or academics after feedback conversations. The Lithuanian system and its main players I had already come to know during my work as consultant for the OECD from August 2014 until January 2016. In this scope, we conducted a Public Governance Review of Lithuania, and visited the country to interview a broad variety of relevant stakeholders there. Furthermore, I had provided background research and drafted chapters on the history and context of the public administration, as well as the development of the performance management system of the country. Hence, I had already been equipped with substantial background knowledge before I carried out the interviews for my own research in Lithuania. Therefore, I spent only two weeks on conducting my interviews in Vilnius. This asked for a different approach. Based on my extensive literature and document research on the Lithuanian performance management system before visiting the country, I identified key actors (i.e. ministries, agencies, and other institutions), and consequently contacted the Office of the Government, the coordinating institution of the performance management system, to provide me with names of the responsible persons in the respective ministries and agencies. However, during the interviews I also came to know about additional actors or potentially interesting perspectives, so also in Lithuania I found supplementary interview partners via the snowball principle.

In addition, I analyse relevant Finnish, Lithuanian, and international (policy) documents, which provide valuable background and context information. These include sources from the Finnish and Lithuanian governments and administrations, but also international sources, such as the OECD’s Public Governance Review of Finland of 2010 (OECD, 2010), the Governments for the Future project report (Ministry of Finance, 2013a), and the OECD’s Public Governance Review of Lithuania (OECD, 2015a). A systematic comparative document analysis complementary to the interviews is not provided, though. This is mainly because I specifically wanted to gain knowledge about how the systems work and how they are perceived in practice, as opposed to on paper. In particular, learning and accountability are concepts whose *de facto* application might differ substantially from *de jure* regulations. For instance, it is hardly imaginable that an organisation would say it is against learning. In reality, however, it might be that perceived pressure to perform, blame avoidance strategies, or (perceived) threat of serious sanctions – as described in the theoretical part of this thesis – are in fact inhibiting genuine learning to take place. Hence, my intention was to focus on how the systems work *in practice*.

For Finland, I base my analysis on 13 interviews (in 11 different departments, with one follow-up meeting and two interviews conducted separately with two respondents of the

same organisation) carried out between April and May 2014 with current and former civil servants of the Finnish Government, which included the following: Representatives of the Prime Minister's Office, the Ministry of Finance, the Ministry of Justice, the Ministry of Transport and Communication, the State Treasury, the Finnish Customs agency, the National Audit Office, the Auditing Committee of the Finnish Parliament, the Finnish Institute for Public Management (HAUS), and Broad Scope (a private management consultancy). These interviewees represent the different actors in the Finnish performance management system: The developers and coordinators, the strategic executors, the operational executors, the auditors, and the internal and external consultants.

Similarly, for Lithuania I conducted 10 interviews in April 2015 with current and former Lithuanian civil servants. As this case study has been started after the Finnish one, I have deliberately chosen the institutions that perform comparable functions as their Finnish counterparts. The interviewees work in the Prime Minister's Office, the Ministry of Finance, the Ministry of Economy, the Ministry of Social Security and Labour, the Ministry of Education, the Ministry of the Interior, the Lithuanian Customs Agency, the Sunset Commission (the Commission for the Improvement of State Administration, under the coordination of the Office of the Government), the National Audit Office, and ESTEP, an independent, private consultancy. In addition, informal follow-up and validation meetings with academics and stakeholders have taken place.

The interviews were thematically semi-structured. I selected a few key issues to present the general opinion of the interviewees. However, most of the interviewees started telling their stories immediately themselves, so that it was not necessary in all cases to ask about these topics explicitly.

The selected issues for the interviews were as follows:

- Recent changes to the performance management system & their origin
- Nature of the performance management system
- Current challenges & vision / solution(s) for the future
- Purpose of the system
- Design of the performance management system
- Use of performance information
- Accountability – Reporting of performance (indicators)
- Learning forums / intra- & inter-organisational relationships
- Compatibility of accountability and learning purposes

The interviews were recorded and transcribed, as this guarantees “the best way to capture the language and phrasing of the interviewees, thus the best way to capture their perspectives” (Luton, 2010, p. 41). However, and especially because some respondents appeared to feel a somewhat intimidated by the recording device, I emphasised that their confidentiality and anonymity would be warranted. For similar reasons, the empirical part does not refer to the respondents' role in their organisations.

After transcribing the interviews, I subsequently coded the information according to the interview issues. Nevertheless, additional themes also emerged when analysing the interviews, so this research is both theory-guided (deductive) as well as data-guided (inductive). Furthermore, several interview issues are of such nature that they often cannot be phrased explicitly. On the other hand, if they had been, this could have led to answers that might reflect social desirability. These include, for example, the extensive topic of learning and learning forums. First, learning forums are an academic term with which respondents might not be familiar. Hence, I asked about “(learning) networks”, “dialogue and reflection moments” and about the characteristics of these forums. Are participants considered to be on equal level? Do they know each other? Can they (and do they) discuss issues critically? How often do they meet? What happens with the results or discussions after these meetings? Similarly, learning is also an abstract concept. When coding, I used my working definition as cited in section 3.3: “the acquisition of knowledge or skills through study, experience, or being taught”. In some cases respondents explicitly told about “learning experiences”; in other cases it was linked to improvement or gaining knowledge. For accountability, coding was slightly easier, as issues such as control and reporting practices are talked about explicitly. In addition, blame-avoidance and gaming and cheating issues became known often without explicitly asking about it from my part.

In sum, I examine what the interviewees’ stories tell us about the perception of the performance management system across the different levels of players, and link it to my analytical model of the use of performance information for learning and accountability to determine whether accountability- or learning-elements prevail in the current system, and what the implications of either are for the functioning of the system.

I opted for a thematic approach to the analysis of the data, which offers the opportunity to examine common issues from different perspectives. This method underlines the multi-dimensionality of the system, and the approach of the actors towards it. The empirical part (chapter 6) is structured according to the main concepts found in the theoretical part. These include:

- The purpose of the performance management system,
This section looks at what respondents consider the vision for and the purpose of the performance management system to be. Here already some tendencies regarding accountability- or learning-based systems can be found.
- its design,
Here, the actual design of the system is analysed. How is it developed and implemented, who is involved, how are targets set, what is the character of the system like?
- the use of performance information,
Questions addressed in this section include the following: What is the purpose of performance indicators? What kind of performance information is used? How is it used? Why is it used? Who uses it?

- reporting of performance information,
This section looks at the how performance results are reported, to whom, and whether there are any sanctions or rewards for negative/positive performance.
- the structural and cultural context, and
This is the most implicit part. Respondents reflected on their opinion about the indicator system and the character of the system as a whole. Also issues such as a control- or trust-based culture came to the fore. In particular, in Lithuania also the cultural and structural heritage of the past was mentioned.
- learning forums.
Here respondents told about any formal or informal networks where any form of reflection or “lesson drawing” takes place, and if so, how they can be characterised. For example, do they have a formal or informal character? Who is involved? What is the relationship among the participants? How is the setting? What is the basis of the dialogue? (Quantitative and/or experiential knowledge, basic assumptions, organisational goals)

In form of a general assessment of the systems, the (in-) compatibility of accountability and learning in the Finnish and Lithuanian performance management systems are discussed in sections 6.2 and 6.4, respectively.

Finally yet importantly, a few words on the individual vs. the organisational learning perspective merit some attention. The reader might ask, ‘what can the researcher tell about organisational learning when she interviewed individuals?’ As outlined in section 3.3.1, I consider individual and organisational learning to be closely intertwined i.e. individuals are the founding pillars of organisations and, as such, organisational learning primarily occurs through individual learning. At the same time, organisational learning is not just the sum of individual learning. Hence, organisations can “know” more, but also less than their individual members can. To gain an understanding of the organisational perspective (through, and although, interviewing individuals), the questions and topics for the interviews were deliberately chosen to reflect on the perception on the organisational level.

4.3. Reflections on the approach chosen – from reliability, validity, and generalisability to trustworthiness and exemplary knowledge

Conducting case studies does not occur without (supposed) drawbacks. Issues such as reliability, validity, and generalisability of qualitative case studies are concerns qualitative researchers are often confronted with and criticised about. However, these concerns mainly result from the quantitative-qualitative divide, with the former, positivist studies being considered as “objective, hard research”, while qualitative researchers are often confronted with being subjective, selective, and not able to generalise.

4.3.1 From reliability and validity to trustworthiness

The concerns about validity and reliability in qualitative research are based on the felt need to make qualitative and quantitative research “equal”. Generally, validity refers to whether an indicator really measures what it is supposed to measure. Reliability then addresses “replicability” in the sense that applying the same methods under the same conditions will deliver the same results (Dodge, Ospina & Foldy, 2005, p. 295). This makes surely sense in a research-setting where a large amount of data is statistically analysed with the aim to draw general conclusions about relationships between variables.

However, the basic underlying assumptions of qualitative research are by nature incompatible with these conceptualisations of reliability and validity. The aim of qualitative research is to shed light on a specific issue, examine it in its multi-dimensionality and complexity, and through in-depth analysis dig out the deeper meanings, causes, and explanations of the issue under study in its specific context and time. It is thus not about getting to know the “universal truth” about the research object, as “truth” is always something subjective and context-dependent. When discussing the “truth” of a specific issue, it is actually the “individual truth”, what appears true to the people involved, to which we are referring. So in fact, it is about authenticity, or credibility, both to the people connected with the issue analysed and to the readers, rather than about validity in the positivist sense of the word. The question one has to ask to assess credibility is whether the story, the narrative is plausible (“true”) for both internal and external audiences (Dodge et al., 2005, p. 295; Lincoln & Guba, 1985).

Also according to Lincoln and Guba (1985) we have to consider the trustworthiness of a case study if we want to evaluate its value. Trustworthiness consists of credibility, transferability, dependability, and confirmability.

Credibility is about authenticity, both towards the people connected with the issue analysed and towards the readers. Transferability refers to the extent to which the findings could be applied in other contexts. Dependability and confirmability relate to whether the research process sounds “fair, unbiased or coherent” to external audiences. As one of the premises of qualitative research is that it is dependent on the context (time, place, respondents, etc.), qualitative researchers do by no means claim that their studies are replicable. Quite on the contrary, they are certain about the fact that qualitative studies cannot offer this kind of reliability. However, what qualitative studies can offer is dependability and confirmability. Does the research process sound “fair, unbiased or coherent” to external audiences? The key here is transparency about data collection and analysis methods (Dodge et al., 2005).

I have put my research results to the test of trustworthiness by organising a workshop in Tampere (Finland) which was attended by scholars and civil servants. I presented some preliminary findings – the draft outline of my story – to assess whether this story sounds reasonable, coherent, and unbiased to the relevant audiences. In Lithuania, I validated my preliminary findings in informal conversations with government representatives and

outside consultants and academics, and I sent draft versions of my results to relevant stakeholders, and received valuable feedback, which I incorporated in the write-up.

Furthermore, I increased credibility by triangulating data sources and relying on multiple informants (Dodge et al., 2005); I used interview material, official government documents, as well as international studies on the Finnish and Lithuanian performance management systems. The interviewees were people working in different organisations of the respective public administrations, as well as employees of in-house and independent consultancies.

4.3.2 From general to exemplary knowledge (*phronesis*)

Case studies often have to defend themselves against the claim that ‘the results might be interesting, but unfortunately, not generalizable’. Similar to reliability and validity, the claim for generalisability originates from the natural scientific method. However, several authors have questioned this assumption and called for a “rethinking” of the case study method (Flyvbjerg, 2006; Thomas, 2011b). Thomas (2011b, p. 33) explains that “special generalisation in social science is an illusory goal: it is not possible in *any kind* of social inquiry.” As such, accusations of not being able to generalise from case studies lose ground. As a matter of fact, Flyvbjerg (2006) clarifies that case studies can be used for generalising knowledge, by referring to (among others) Darwin’s, Newton’s, or Galileo’s “research findings”, which were not based on large-scale, quantitative analyses, but on single case studies. Nevertheless, he claims that “formal generalisation is overvalued as a source of scientific development, whereas ‘the force of example’ is underestimated.” (Flyvbjerg, 2006, p. 228).

Thus, rather than focusing on general, context-independent (and generalizable) knowledge, in the social sciences we should in particular focus on exemplary knowledge, i.e. concrete, context-dependent knowledge (Flyvbjerg, 2001, 2006; Flyvbjerg, Landman & Schram, 2012a; Thomas, 2010, 2011b). Exemplary knowledge is to be understood as “a particular representation given in context and understood in that context. ... It is interpretable only in the context of one’s own experience.” (Thomas, 2011b, p. 31). Exemplary knowledge is gained through *phronesis*, an approach which goes back to the Greek philosopher Aristotle. Flyvbjerg (2001, p. 2) explains that *phronesis* “goes beyond both analytical, scientific knowledge (*episteme*) and technical knowledge or know-how (*techne*) and involves judgments and decisions made in the manner of a virtuoso social and political actor.” As such, at the heart of *phronesis* is “the discussion of values and interests, which is the prerequisite for an enlightened political, economic, and cultural development in any society” (Flyvbjerg, 2001, p. 3). Consequently, if social science really wants to matter, it should be based on “context, judgement, and practical knowledge” (Flyvbjerg, 2001).

Flyvbjerg (2001, p. 60) suggests four value-rational questions as a starting point for phronetic social research:

- 1) Where are we going?
- 2) Is this desirable?

3) What should be done?

4) Who gains and who loses; by which mechanisms of power?

Answers to these questions are kept in mind while conducting the analysis, and are addressed particularly in the concluding chapter.

5. MAPPING OUT THE TERRAIN: FINLAND AND LITHUANIA IN CONTEXT

*A good system shortens the road to the goal.
~ Orison Swett Marden*

This chapter aims to introduce the context of the Finnish and Lithuanian performance management systems. The two systems in their respective contexts are presented in turn. Each case is described in its politico-administrative context, before the concrete performance management system, its development, and recent changes and challenges are introduced. This descriptive part is based on academic literature, as well as country sources (such as reports from the respective ministries) and OECD reports.

5.1. A politico-administrative description of Finland

Finland is a unitary state with highly autonomous local authorities. It is furthermore a unicameral parliamentary democracy, and generally considered a consensual democracy. The latter is notable by fairly stable coalition governments (Pollitt & Bouckaert, 2011).

Finland's state central administration is divided into three levels: the national, regional, and the local level. The national level consists of the Parliament, the President, and the Government with the Prime Minister and 12 ministries. These ministries work quite independently from each other in preparing and steering the implementation of policy decisions. Furthermore, they steer the agencies, mostly by means of performance agreements. About a hundred agencies and public bodies employ about 20,000 people. Agencies work quite independently under the respective ministries and are responsible for policy implementation and service production, and also fulfil advisory functions (Ministry of Finance, 2010).

A distinctive feature of the Finnish public administration is the independence and self-government of the municipalities. In comparison to other EU countries, municipalities have an extraordinary variety of responsibilities. They organise the majority of public services directed to citizens, which, according to different studies, are used widely among citizens, and are perceived positively (Ministry of Finance, 2010). However, as the present study focuses only on the central level, an elaboration of the local dynamics goes beyond the scope of this dissertation.

Regarding cultural elements, apart from being considered fairly homogenous and egalitarian (Pollitt & Bouckaert, 2011), generally a high-trust environment is attributed to Finland. According to the latest available Eurobarometer results, 67 % of the respondents "tend to trust" the Finnish Parliament, and 60 % "tend to trust" the Finnish Government. This is twice as much as the EU average, which is located at 31 % for both

the national government and the national parliament (European Commission, 2015a). A comparison with earlier Eurobarometer results furthermore shows a slight increase in trust (e.g. in spring 2013, the results showed 61 % trust in the Parliament, and 54 % in the Government) (European Commission, 2013a).

5.2. The Finnish performance management system – “an agreement-based control system”

In the scope of the NPM movement in many western industrial countries, also Finland started to modernise its administrative system. Plans for the introduction of a performance management system date back to 1988, and by 1995 the system was fully implemented. Against the background of the “exceptionally deep recession” in Finland in the early 1990s, this reform was aimed at increasing “the freedom of action and accountability of government agencies and institutions in order to gain better services and more efficient use of resources” (Ministry of Finance, 2010, p. 41). To achieve this aim, public service functions were converted into public enterprises, and management principles were introduced on all levels of government. An important aspect was the relationship between ministries and agencies. While agencies are granted lump sums from the ministries and independence in the way the appropriations are spent, in return the agencies are expected to commit themselves to performance targets agreed on by means of performance agreements (Ministry of Finance, 2010, p. 41).

The performance management system has been developed further since. Incentives and inspiration for reforms were received both by internal reviews (e.g. performance reports, hearings in the Parliament, etc.) and by outside reviews (the Governance Review by the OECD in 2010 played a major role in defining challenges and opportunities for the Finnish system). In addition, the Finnish Institute for Public Management, HAUS, which serves as an in-house consultancy in close connection to the Ministry of Finance, contributed largely to the evaluation and reform of the performance management system in Finland.

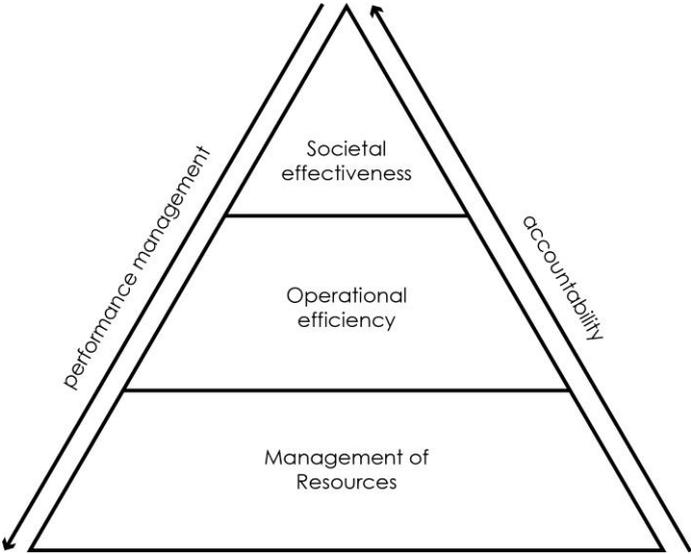
Nowadays, the Finnish performance management system is officially described as an “agreement-based interactive control model” (Salminen & Viitala, 2006, p. 7). The key document in the performance management process is the annual budget, which contains performance targets of all levels on a general level. Furthermore, the “agreement” element consists of the performance agreements between ministries and their respective agencies. These documents are at the core of the Finnish performance management system. In these agreements, more specific performance targets are laid down. While not legally binding, the performance agreements constitute a “morally binding document of mutual understanding” (Ministry of Finance, 2013b, p. 5). The performance targets are agreed upon during a dialogue and negotiation phase, in which representatives of the respective ministries and agencies take part.

The “control” element of the Finnish performance management system and its relation to performance is generally presented as “prism” (see Figure 8). Performance management

is organised top-down. The Government sets up its Government Programme and an accompanying Government Strategy Document. These documents state the Government’s strategy and focus for the Government’s term of office of four years, and include the societal effectiveness targets. From there, target-setting cascades down to strategic and policy targets (mostly located in the ministries), and subsequently operational targets (at the agency level). Each ministry derives its own targets and the targets for its agencies from these documents. Performance targets for agencies include both basic operation targets and development measures. In the majority of the cases, these performance targets are defined by means of (quantitative) indicators, with only a few qualitative measures (Ministry of Finance, 2010, p. 45). Performance agreements are usually agreed on for one year, although lately (and in the scope of the most recent performance management reform) an increasing number of ministries have decided to set up performance agreements for several years.

Accountability, naturally, goes the opposite direction. The reporting goes bottom-up: Agencies have to report to the respective ministries about their performance during and at the end of the year. Ministries report to the Government and Parliament. Outside this chain, there is the National Audit Office (*Valtiokonttori*), which regularly audits the performance of ministries and agencies. This audit office has been working independently in connection with the Parliament since 2001. In addition, also the Parliament has its own Auditing Committee, which mainly audits the Government finances and compliance with the budget (Ministry of Finance, 2010, pp. 93–95).

Figure 8 Performance prism



Source: Ministry of Finance (2013b)

In sum, the main actors and responsibilities of the Finnish performance management system are the following:

- The *Prime Minister's Office* and the *Ministry of Finance* are responsible for the development and the coordination of the performance management system in general. They lay down the main (strategic) targets for the whole Government and public administration, respectively.
- The *ministries* set strategic targets for policies under their competences, and negotiate performance agreements with their agencies. They are responsible for steering the policy planning and implementation.
- The *agencies*, which are linked to the performance management system by the performance agreements with the ministries, are mainly responsible for operational tasks.
- The *National Audit Office* and the *Auditing Committee* of the Parliament audit all actors in the performance management system.
- Finally, *in-house* and *outside consultants* (like HAUS or the OECD) play an important role in evaluating and developing the performance management system.

5.3. Current situation and challenges

Both the agreement- and the control-elements of the Finnish performance management system have been criticised for various reasons: Performance targets and output measures appear to have been “biased and distorting”. This, in turn, makes the system more rigid than flexible, and affects the “strategic development practices” within the system. Regarding the agreement elements, performance negotiations are not characterised by genuine dialogue or negotiation, and instead pursue a top-down approach. Finally, the reporting requirements led to agencies providing a considerable amount of reports, which, however, lack information about “operational effectiveness or operational performance” (Ministry of Finance, 2006, pp. 21–22).

In its Public Governance Review of Finland, the OECD found that Finland has to invest more energy and resources in strengthening a shared vision and horizontal collaboration. Based on the fairly homogenous public service, emphasis is put on informal and cross-sectional networks. In fact, informal decision-making processes are considered to be at the core of the Finnish public administration. Nevertheless, more investment in leadership and clear communication is needed to implement the shared vision and to make decision-making more evidence-based. Furthermore, while the performance management system is considered to be well structured and focused on outcomes, it lacks flexibility and agility (OECD, 2010).

In addition, the National Audit Office and the Parliamentary Audit Committee have reported several shortfalls of the performance management system. These include a lack of information-based policy-making in the social and the health sectors; a lack of clear

indicators which link to the government targets; a lack of quality of the performance indicators; insufficient reporting about the achievement of the targets; and a lack of consequences for failure to meet the targets (OECD, 2010). Against this background, the Ministry of Finance has initiated a performance management reform which rests on four pillars (Ministry of Finance, 2012):

- 1) Making the system more strategic,
- 2) Reducing the administrative burden,
- 3) Making the system more horizontal, and
- 4) Making the system more integrated.

The first point refers in particular to the goals and targets included in the performance agreements. The performance agreements are to include less, but more strategic targets that are to be mutually discussed among agencies and ministries. Second, the reduction of the administrative burden is to be achieved by making the system “lighter”, and thus easier to manage. The idea is to set up the performance agreements for a four years’ term, in line with the Government Programme, rather than for one year, as has been the case in most ministries. This is to ensure a better connection, cascading down from the goals and targets in the Government Programme to the ministries and from there to the agencies. In addition, the follow-up to the fulfilment of the targets is to be improved. Next, by encouraging ministries to work more closely together and across sections, the system would become more horizontal. Fourth, to make the system more integrated and uniform, a common performance agreement template has been developed, which ensures that the same information is included in all performance agreements across ministries and agencies. In addition to these four main points, also the introduction of shared common goals is envisaged. These should be derived from the Government Programme. It is expected that these goals will offer the agencies the opportunity to link their actions to the “bigger picture” as well (Ministry of Finance, 2012). This performance management reform also played an important role in the interviews with the Finnish public administration employees, and as such, I return to it in the next chapter.

Another influential analysis of the current situation and challenges and opportunities has been provided by the Governments of the Future project in 2013 (Ministry of Finance, 2013a). To be better prepared for the future, Finland wants to move “from detailed steering to understanding systems” (Prime Minister’s Office & Ministry of Finance, 2013, p. 19). Table 7 below summarises the main differences between the old and the future governance approach, as proposed by the Prime Minister’s Office and the Ministry of Finance.

Table 7 Main differences in focus between the old governance approach and the future approach

Focus of old governance	Focus of new model
Vertical accountability	Master future challenges
Results	Sense-making
Control	Trust
Monitoring and surveillance	Learning and reflection
Competition	Collaboration
Yearly focus	Longer term
Reacting to symptoms	Understanding of the underlying reasons
Linear causal connection	Complexity
Input	Impact
Detailed steering	Understanding systems
Not rocking the boat	Controlled risk-taking
Outer motivation	Inner motivation
Unit cost	Total cost
Citizens as objects	Citizens as subjects
Reports	Experiments

Source: Prime Minister's Office & Ministry of Finance (2013, p. 20)

Interestingly for the present research, the new model focuses in particular on “learning elements”, as discussed in the theoretical part of this dissertation. Sense-making, trust, learning and reflection, but also collaboration, inner motivation, understanding systems, and recognising complexity receive attention. To foster an atmosphere of trust, investment in open and transparent learning dialogues is recommended. Furthermore, the report recognises that the “successfulness of a performance system depends on to what extent various actors have been able to participate and cooperate within the process of setting targets and particularly in assessing the results and outcomes.” (Prime Minister’s Office & Ministry of Finance, 2013, p. 39). Hence, improving simply the quantitative measurement (i.e. single-loop learning, as discussed in section 3.3.2) is not enough. More investment in double-loop learning is needed.

To sum up, Finland has some 20 years’ experience in performance management. While initially designed and implemented against the background of, and inspired by, the NPM-movement in western industrial countries, the underlying rationale and principles of NPM have been critically reviewed. These critical reviews have also been supported by outside sources, such as the OECD or the participants of the Governments for the Future project. Thus, Finland shows great willingness to reflect upon its practices, learn from

and with others, and adjust its system and practices where necessary. Based on these evaluations and recommendations, recent reform efforts have striven to make the performance management system more agile and flexible, making it better prepared for a fast-changing environment. Furthermore, dialogue among all involved stakeholders, horizontal cooperation, and learning and reflection are considered important elements to achieve these aims. The egalitarian, homogenous culture, paired with a relative high-trust environment, appear to be promising factors for success in the Finnish case. Hence, Finland can be considered a most likely case for a performance management system that is based on learning. After presenting the Lithuanian context in the following subsection, the empirical chapters empirically investigate this premise.

5.4. A politico-administrative description of Lithuania

Similar to Finland, Lithuania is a unitary state with highly autonomous local authorities. It is furthermore a unicameral parliamentary democracy (OECD, 2015a).

Lithuania's State central administration is divided into two levels – the national and the local level. The national level consists of the Parliament (the *Seimas*), the President, and the Government with the Prime Minister and 14 ministries. Prior to 2009, there had also been a territorial level (the counties). However, in the scope of several reforms (the most influential one being the “county management reform” of 2009), the administration was restructured. Functions which had formerly been performed by the county administrations were transferred to either the municipality or the national level to make the system more efficient, bring it closer to the citizens, and reduce administrative burden (OECD, 2015a; Sustainable Governance Indicators (SGI), 2014).

The 14 ministries work independently from each other and lack horizontal coordination. This is seen as a legacy of the communist history of the country. Due to the separation of state and communist party, at the end of the communist regime Lithuania was faced with an underperforming state structure. Civil servants were specialists, there was no uniform civil service, and as such, internal coherence was low. The wish to restrain powers of individual central government institutions to avoid one particular politician to become the authoritative power led to the creation of a dominant bureaucracy (Nakrošis, 2001).

Regarding cultural elements, similar to Finland, Lithuania is also described as fairly homogenous. However, quite the contrary to Finland, a low-trust environment is attributed to Lithuania. According to the latest available Eurobarometer results, only 16 % of the respondents “tend to trust” the Lithuanian Parliament, and 34 % “tend to trust” the Lithuanian Government. (The EU average is located at 31 % for both the national government and the national parliament.) Trust in EU institutions is located at 68 %, as opposed to 40 % of the EU average. This is surprising, as it shows that Lithuanians tend to trust the EU institutions more than their own national institutions (European Commission, 2015b). However, a comparison with earlier Eurobarometer results shows a slight increase in trust (e.g. in spring 2013, the results showed 13 %

trust in the Parliament, and 27 % in the Government) (European Commission, 2013b). These results are in stark contrast to the results for Finland.

In addition, the Lithuanian public administration is characterised by a strong legalistic and hierarchical nature (Nakrošis, 2001; Pivoras, 2013). Pivoras (2013) explains this referring to three different historical layers that shaped the Lithuanian administrative culture by emphasising the legal dimension of public administration: “the heritage of independent statehood from the interwar period ..., the Soviet legacy, and the constructed or imagined European administrative tradition”. Paradoxically, the transposition of EU legislation into national law, which has been a prerequisite for EU accession and integration of Lithuania, has re-enforced this legalistic dimension rather than countering it. (At the same time, it might not be too surprising, giving the EU’s extensive history as “indicator bureaucracy” – examples are the Stability and Growth Pact, the Lisbon Process, EU2020, and the abundant list of indicators required for receiving funding in the scope of the European Structural Funds.) Furthermore, Pivoras (2013) also links the lack of trust in government and parliament to the Soviet legacy. In the following section, we see this legacy also in the performance management system.

5.5. The Lithuanian performance management system – “a formalised system for performance monitoring and reporting of results”

Performance management in Lithuania officially carries the name “strategic management”. It is defined as a “formalised system, ensuring the realisation of the whole strategic management process, during which performance directions and ways are defined [...] for achieving planned aims as well as performance monitoring and reporting of results.” (Nakrošis, 2008). As such, we can confidently talk about “performance management” here for reasons of comprehensibility and comparability, as the system consists of all stages of performance management (planning, measuring, monitoring, evaluating) (Nakrošis, 2008).

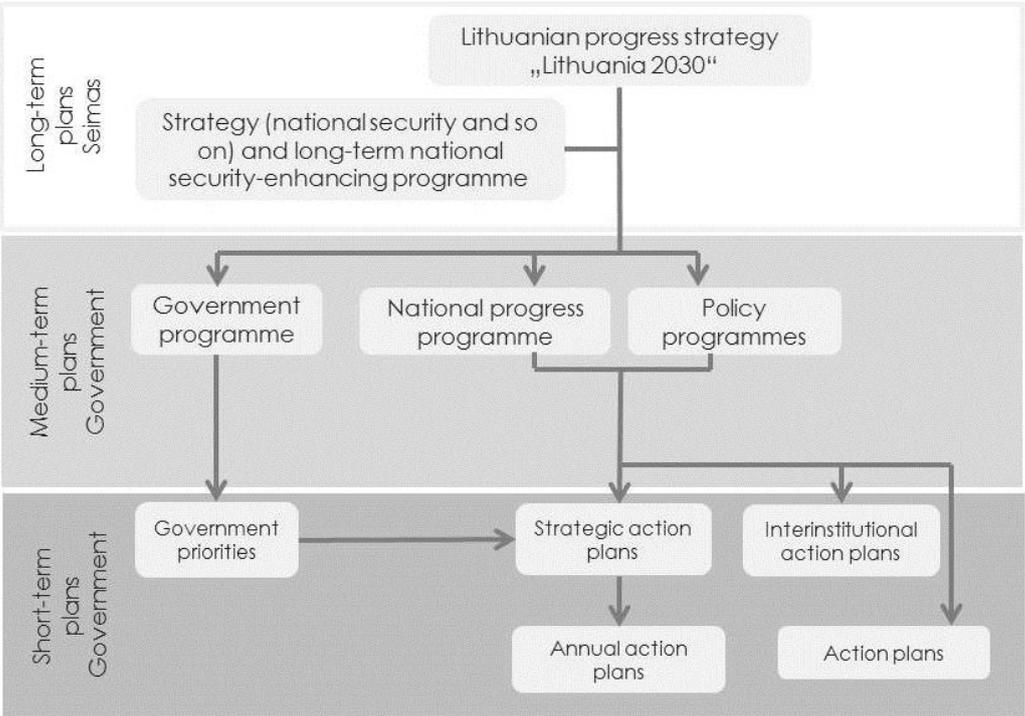
Contrary to Finland, where no specific legal documents exist for the regulation of the performance management, in Lithuania the performance management system is legally documented in a variety of documents. The most encompassing document is the Law on Public Administration, which focuses particularly on monitoring performance and changes in the public administration (Republic of Lithuania, 1999). More specifically targeted towards performance management is the Strategic Planning Methodology, in which the major guidelines are written down (Government of the Republic of Lithuania, 2002). To help institutions set their targets, there is furthermore a methodological guideline for formulating and analysing performance indicators (Office of the Prime Minister of The Republic of Lithuania, 2011).

Performance management has been introduced in Lithuania in the year 2000 in the scope of the preparations for EU and NATO membership (Nakrošis, 2008; OECD, 2015a). It is in essence a top-down system, with the Office of the Government and the

Ministry of Finance being responsible for the overall coordination. Contrary to Finland, Lithuania does not work with performance agreements among ministries and agencies. However, a detailed and binding strategic action plan system is in place (see Figure 9). These strategic action plans are divided between long-term, medium-term, and short-term planning documents. The strategic action plans contain indicators for products, outputs, and outcomes. Information about these indicators, as well as all documents related to strategic planning, are fed into the centralised Monitoring Information System (SIS) on quarterly and yearly basis.

Since its introduction, the performance management system has been reformed in various steps (For example, the “Improvement of Performance-based management” (VORT) project from 2009-2012, and the ongoing “Public Governance Improvement Programme – 2012-2020”). The latest, ongoing reforms have focused mostly on making the system more efficient and results-oriented. In addition, a stronger emphasis on monitoring and evaluating performance information is envisaged.

Figure 9 Strategic planning system in Lithuania



Source: Office of the Government (provided by e-mail, April 2015)

While not having a fully-fledged performance-budgeting system in place, the strategic planning process and the annual budget process are closely interlinked.

In brief, the main actors in the performance management system of Lithuania are the following:

- The *Office of the Government* is responsible for coordinating the design, improvement and implementation process of the strategic planning documents and assuring the quality of these documents, together with the *Ministry of Finance*, which is primarily in charge for the appropriations in these plans. In addition, the Office of the Government is in charge of elaborating cross-cutting strategic planning documents, and monitoring the performance results of strategic plans (National Development Programme, Government Priorities). The Office of the Government also drafts the annual report of the government, and coordinates the work of the Commission for the Improvement of Public Management, the so-called *Sunset Commission*.
- The *Ministry of the Interior* is in charge of coordinating the implementation of the Public Governance Improvement Programme for 2012-2020, which aims, among others, at improving “result-oriented and evidence-based management” (Government of the Republic of Lithuania, 2012).
- The *line ministries* and *agencies* are responsible for drafting and implementing their respective strategic plans. They also report on their performance to the Office of the Government, respectively their line ministries.
- The *National Audit Office* scrutinises financial compliance and implementation of strategic action plans. Performance audits are undertaken on a regular basis.
- Cooperation with *in- or outside consultants* is less frequent in Lithuania than in Finland. Nevertheless, a recent Public Governance Review led by the OECD and a training programme implemented by ESTEP, an outside consultancy, serve as illustrative examples for efforts to develop and improve the existing performance management system. Furthermore, the comprehensive “Results-based Management Improvement (VORT)” project has been implemented together with outside consultancies (Ernest & Young, a group from the Public Policy and Management Institute – PPMI – and ESTEP). This project was aimed at overhauling existing practices and introducing performance management components into the planning and budgeting system.

5.6. Current situation and challenges

Despite several reforms to make the Lithuanian government more result-oriented, it is questioned whether performance management actually contributes to making government and public services more effective and efficient. While substantial performance information is available, it appears to be underutilised by public officials (Nakrošis, 2008). The abundance of legal regulations makes civil servants focus more on the compliance with the rules at the expense of a focus on actual implementation of programmes (Nakrošis, 2001).

Several reforms have aimed at overcoming these deficiencies. One of the most important efforts to reform the performance management system has been the “Improvement of performance-based management” (VORT) project. This project aimed to “improve performance monitoring, accountability, functional analysis and decision impact assessment, and as such performance effectiveness of public administration institutions” (OECD, 2015a, p. 51). In the scope of this project, the Monitoring Information System (SIS) has been introduced and implemented. This system is a single entry point for all performance information of institutions accountable to the government. It also contains performance reports, strategic documents, and links to other performance indicators. It is hoped that this system is used to support decision-making (OECD, 2015a). Performance information is fed into the system on a quarterly basis to ensure the timeliness of the information. Achievement of targets is demonstrated visibly by the use of a “traffic light system”. However, the number of performance indicators increased rapidly, exceeding more than 3 000 indicators in 2013. In turn, this increased the administrative burden, and asked immense reporting and assessment capacities and time investment from the staff (OECD, 2015a; Rimkute, 2014).

Against this background, and to reduce the administrative burden and make the system more actionable, the “Improvement of Strategic Planning Documents Performance Monitoring and Evaluation Processes” (RESST) project (2013-2015) was run by the Office of the Government. One of the goals of the project was to reduce the number of performance indicators and provide methodological assistance to ministries in order to develop a reasonable and manageable set of performance indicators (information provided by the Office of the Government, August 2015). Furthermore, top managers were coached to increase their capacities to use performance information, and civil servants received trainings on how to improve the quality of performance indicators, as well as their analysis. In addition, staff was trained to better feed performance information into the budget negotiations and government priorities (OECD, 2015a).

To sum up, Lithuania has introduced a comprehensive performance management system at the beginning of the 2000’s. Implementation was triggered and influenced by accession to EU and NATO in 2004, and against the background of a communist past. As such, Lithuania has had to face a number of challenges, such as a low trust environment, and a hierarchical and legalist culture, which led to more focus on compliance with the rules rather than on policy implementation. This resulted in a hierarchical performance management system with extensive reporting requirements. Ironically, it was also during the EU accession period, when the legalistic administrative culture and the strong emphasis on accountability were reinforced. This was due to the large amount of EU legislation which had to be transposed into Lithuanian law (Pivoras, 2013). Hence, there is an abundance of performance information required and available; however, both its quality and actual use seem limited. The focus on following the rules is stronger than on implementing them, and external and internal control is strong. These reasons provide ground for the premise that in Lithuania we are more likely to encounter elements of an accountability system, as opposed to elements of learning systems.

6. IN THE FIELD: THE FINNISH AND THE LITHUANIAN SYSTEMS ANALYSED

*Sometimes we simply have to keep our eyes open
and look carefully at individual cases – not in
the hope of proving anything, but rather in the
hope of learning something!
~ Hans Eysenk (1976)*

This chapter provides the empirical analysis of the performance management systems in Finland and Lithuania. It is based primarily on interview material, supplemented with an analysis of relevant documents to clarify certain points where deemed necessary. First, the findings of the Finnish case are presented, before the Lithuanian findings are elaborated. The results of both cases are outlined according to the main themes used during the interviews i.e. the purpose, the design, and the actual use of performance information, followed by the evaluation and assessment of performance results. To capture the context, the culture and structure in which the performance management systems are embedded, are analysed subsequently. Thereafter, the existence and mechanisms of learning forums in the respective cases are examined. Each case analysis ends with an evaluation regarding the underlying assumption of the likelihood of being a case for learning, respectively accountability. The final subsection compares the two cases, concludes, and presents the implications of the findings.

6.1. Use of performance information in Finland

6.1.1 On the purpose of the performance management system

I asked all the participants what they considered the purpose of the performance management system. The variety of answers received to this question reveals the multi-dimensionality and complexity of the system.

The answers can be divided into three main themes. The first theme circulates around the functions of accountability, steering, and control. The performance management system is considered to guide policy actions, with performance targets serving as “checkpoints” to see whether these actions go into the right direction. Furthermore, performance targets are to provide guidance for knowing what someone is supposed to do. In addition, resource allocation and effectiveness are considered important purposes of the performance management system.

The second theme is about more abstract and broader purposes of the performance management system: to reach a “*better society*”, tackle societal problems, and to provide true and fair information. In this sense, the public sector is seen as having a distinct competitive advantage opposed to the private sector, as one respondent (R1) claims:

“[You] have kind of a competitive ... advantage there, that there’s no business organisation in Finland or in some other country which has kind of a mission which the public sector has, you know, to make the society better, and so forth...”

The third theme centres on a critical view of the performance management system. The system is considered to have grown out of hand, and thus to have *“become a system by itself”* (R2). In a similar vein, several interviewees complained about the lack of clarity of the purpose in general. Performance targets are considered to exist because the administration *“has to have them”* (R3), rather than to pursue a specific purpose. In addition to these criticisms, also the general NPM doctrine is challenged by admitting that copying strategies from the private sector and applying them to the public sector does not seem to be satisfactory.

In particular the “narrow” focus on performance targets at the expense of considering the context or the *“big picture”* (R1, R4), as frequently voiced by the interviewees, is criticised extensively. While performance targets are considered necessary and good to reflect on effects and results, respondents remain somewhat sceptical about the actual use of these indicators.

The general vision for the performance management system shared by most respondents focuses on four main points, which are also the four main principles according to which the system is to be reformed (see section 5.3). To make the system more horizontal, with ministries working more closely together, one respondent suggests that *“we have to break down [the current] system”* (R1) and get ministries to work with each other. Furthermore, the introduction of shared common goals derived from the Government Programme is expected to offer the agencies the opportunity to link their actions to the *“bigger picture”* as well. The focus on this bigger picture was something that emerged in most interviews. The aim is to reduce the number of targets, while increasing their relevance and focus: fewer targets, but better focused. I return to this point in the following section.

Another aim that was expressed several times was the establishment of a closer link between the budget and performance. This should increase the relevance of the performance assessment and reporting. In this respect, another recommendation is to think of the design of the evaluation while target-setting – rather than *ad hoc* afterwards. This “form follows function” argument is indeed an important aspect in the design of performance management systems; the use of performance information should determine the design of the system (Behn, 2003; Van Dooren et al., 2015).

In summary, the purpose of the system, if recognised at all, revolves mostly around (traditional) ideas about accountability, steering and control, and guidance and feedback. The “soft” side of the performance management system, fostering organisational learning using performance information, is mentioned less explicitly. However, at the same time, there is remarkable criticism about the current (lack of) purpose, and the recognition that the NPM-doctrine might have proven to be less successful than expected.

6.1.2 On the design of the performance management system

As regards the design of the performance management system, the main tendency shared by the interviewees is that the system has a top-down character, and it is considered to be organised in a heavy and bureaucratic way. Decision-making takes place hierarchically, top-down, from the ministries to the agencies. The overall power resides at the Ministry of Finance, which is in charge of the development and coordination of the performance management system. It is mainly the Ministry of Finance, and the other ministries that take the concrete decisions. The respective line Minister has the final word. However, at the same time the Ministry of Finance is criticised for not living up to its tasks and responsibilities, as there appears to be a lack of coordination. Most of the interviewees agree that the system, and most importantly, the performance targets, is set up by cascading down from the Government Programme to the ministries and from there to the agencies. However, in one institution the view is held that the system is actually too much 'bottom-up'. They miss a closer link between performance targets and Government Programme.

Performance agreements are discussed among the heads of the agencies and the responsible directors of the ministries. According to the interviewees, there does not appear to be a uniform way of setting the targets; while some respondents report that the target negotiations are a two-way process, others claim that the targets originate mainly from the Ministries, and the agency is consulted only after the first proposal of the agreement has been presented. Ultimately, however, the head of the Ministry is responsible for performance management.

A remarkable issue is the division between strategic and operational targets. According to the performance prism presented in section 5.2, effectiveness and strategic targets are mainly situated at the ministerial level. However, there are agencies that also wish to see strategic or effectiveness targets on their agendas. Nevertheless, in the performance agreement negotiations, this is either not discussed or, in many cases, dismissed – with important exceptions, of course. This is seen as “proof” for the – in many cases – still hierarchical attitude, especially on the ministerial level. However, not only could this discourage agencies to deliver their results, this also leads to a loss of the view of the wider context and increases the silo-structure of the ministries and their portfolios.

To counter this effect, all respondents (interestingly, also those who promoted this system in the first place) claim that there is an increasing need for cooperation and coordination. Policy issues should replace specific organisations and their responsibilities at the top of the pyramid. In so doing, it would become clear to all participants involved that (most) policy issues cannot be addressed by a single ministry and its agencies alone, explains a respondent. Take for instance the relatively high early retirement age between the 16 and 34 years old. This appears to be caused mainly by mental illnesses. However, to address this issue substantially, all key players have to be brought together, such as different ministries, the social security institute, etc. As such, the respondent telling this example explicitly calls for seeing beyond one's own nose. While, in fact,

many such cross-sectional initiatives have been planned or considered, in practice they often run aground. The respondent (R1) claims:

“There is something mystic in those walls of those institutions because people change, you know, but those basic assumptions, you know, they remain there. ... behind those walls and those corners you have those powers which are kind of maintaining those basic assumptions based on kind of many years ... development and involvement.”

However, not only is the institutional memory so to speak called responsible for a lack of cooperation and innovative thinking. In addition, the behaviour of the managerial level gives room for improvement, according to several interviewees. Ministries persisting on their authority regarding strategic targets and rejecting their agencies' wishes to be more involved in achieving these targets are considered to have a counterproductive effect on the achievement of exactly these targets. In contrast, an organisational culture based on mutual respect and trust is needed. This would also change the role of the leadership (R1):

“If you trust more, you know ... that doesn't mean that you're not, you're not a superior or boss any more, but you are boss in different ways.”

However, also most of the respondents agree that these issues are based on individual experiences and are thus far from being generalizable. There are indeed ministries that involve their agencies to a greater extent than others do, and there are individual managers or management levels signalling a different form of management or leadership. In one ministry, for example, they consider a relationship of trust with their agencies as essential to work together and achieve the set targets. They even say that in essence, this performance management system is *“based on trust”* (R6).

When we examine the vision for the future of the system, the responses can be divided according to two main tendencies.

Some actors call for a more structured, more strategic, top-down, cascading-down approach of the system with a clear distinction between strategic tasks (the ministries' responsibilities) and operational tasks (the agencies). Shared, common goals should establish a cross-sectional connection.

At the same time, not everybody agrees with such a “one size fits all” model, as the different sectors and policies require different actions. Consequently, proponents of the second tendency suggest a more flexible, tailor-made system. This would include better cooperation, dialogue, and cross-sectional work within the public administration. Trust and “true leadership” are also proposed as important qualities of such a system.

6.1.3 On the use of performance information

Closely linked to the questions on how the system is and could or should be designed, is the question of the use of performance information itself. We know now the context in which performance information is (supposed to be) used. Now let us examine what kind of performance information is collected, how this information is consequently made use of, and what this use implies.

The prevailing type of performance information that is used throughout all agencies and ministries is quantitative information, and is considered “easily measurable” and better comparable across ministries and agencies, and easier to feed into the performance information system Netra. Furthermore, it is considered that performance can be assessed more easily when quantitative targets are set.

On the other hand, the critical voices about quantitative performance information did not remain silent. Notably the auditing organisations and “outside” institutions remarked critically on the usefulness of the present performance information. Not only is what is measurable measured, rather than what is sensible, also this type of performance information seems to invite actors to “cheat”, and reverse the original idea of the performance information. Some respondents reported several instances where performance indicators missed the target. One example is an objective that government has to provide “some kind of action” for young unemployed people within three months after their last employment ended to reincorporate them back into the work life. One interviewee explains the perverse effect of performance indicators that measure “actions” rather than “quality” (R3):

“It’s nice objective. You can measure, three months, and yeah, you can count how many you have been able to get to the process. But one action can be that the government agency has agreed with the youngster, yes, you will continue your own approach to get the work place. ... But then you can think, yes, we have had an action, but the action is so-called empty action, there is nothing that the one who came to the agency received. It’s not very effective action. But then we have met our rule for three months, that we have created an action. Then if we report positively, we say that yes, we have had this kind of target, three months, we have met it. But we don’t describe the negative side that we didn’t have enough resources to create relevant and effective measures, that kind of reporting we don’t see.”

Consequently, measuring just for the sake of measuring, because it is prescribed by law, has reached its limits. Several interviewees argue for a shift towards more sensible, useful, and relevant performance information, which might indeed not necessarily be the easiest one to measure. This again links back to the purpose of the performance management system. What is the ultimate goal of the performance management system? Why is the data collected? What kind of data is collected? Moreover, what is it ultimately used for?

As concerns the actual use of performance information, the responses were pessimistic. Several interviewees admit that they do not see the information to be used at all. While

the information appears in the performance reports of the relevant organisations, only rarely is a follow-up analysis provided. Some respondents argue that in particular the performance agreements appear to be more of a “formal step” rather than a binding commitment. Targets are set for the agencies, but in many cases, the final reports do not reflect the agency’s achievement on that set of targets.

The major problem suggested with the low quality of the reporting and the lack of connection between performance indicators and reports is the difficulty of finding adequate indicators. One respondent therefore asks (R3):

“For example, how to evaluate our [organisation’s] performance? If we would have to [have] a certain agreement and monitoring and reporting process, would it be number of reports, like it is now? But that’s not actually good indicator for the quality of our work. So big part of the actions that agencies are performing, they are by definition in such kind of actions that you have, it’s quite difficult to set up good performance indicators. That’s also, I think, one of the reasons.”

To overcome this problem, it is suggested to develop indicators that are relevant and clearly linked to the objective they are supposed to measure. Consequently, this would have to be followed up consistently, and actions should be taken accordingly. This, however, would imply a change in the “mental environment” as well, which would also allow for the reporting of negative performance results. Currently, it seems that underachievement is discussed internally, but not officially reported.

In cases where respondents claimed that they do not see information to be visibly used, I asked them to reflect on the general purpose of collecting performance information and setting up performance indicators. Rather pessimistically, the opinions ranged from “*not clear*” (R2) to “*because we have to have them [i.e., the performance indicators]*” (R3). Again, this seems to relate to the opinion that the performance management system has become out of control, and has developed into a system of its own, instead.

On the other hand, however, some respondents did claim that performance information is used to increase the awareness of the effectiveness of certain actions, especially when discussing performance information in the budget. In particular the strategic targets are considered to be used even “*all the time*” (R6) (though rather not for decision-making), while operational targets are not paid too much attention to. Furthermore, performance information is seen as a helpful guideline to see where an agency or ministry is supposed to go. It can also serve as a tool to demonstrate to the public whether taxpayers’ money is used in an effective way.

One organisation provided the example of how they dealt with the comparatively high sick leave rate among their employees. By analysing the results of a personnel satisfaction study, together with the HR department, this organisation started a “*special project targeted for those offices*” (R7).

They had “preparatory meetings, quite many with these problematic offices, and thereby we little by little more or less identified the main core reasons for the high rates. And then they targeted at those and in those places, the numbers of sick leaves

they have dropped quite a lot. In one place it went down to 1/3. That was of course extreme.” (R7)

This example serves as illustration of how performance results can function in a cooperative action to improve a specific situation. Importantly, the issue in question (the sick leave numbers) were not made public externally, and hence, were only dealt with internally.

In brief, while some organisations do rely on performance information, and use it beyond the mere reporting of it, other organisations do not seem to perceive the benefit of collecting performance information, as it appears not to be used to the extent as expected. In addition, a few respondents mentioned undesired side effects.

6.1.4 On the reporting of performance information

What has become clear from the previous subsection is that it is not only important to collect information, but also to use this information. Performance information is primarily (supposed to be) used for reporting purposes. As stated in section 2.5, the literature distinguishes primarily between hard use and soft use of information. In Finland, mostly the latter is the norm – if at all. This subsection looks more closely into the consequences of performance reporting.

Several respondents point out that performance might not be assessed consistently. They state that no further action results from unmet targets. Consequently, one interviewee claims that *“the accountability system, I don’t think it has ever worked here in Finland.”* (R2) This may be related also to the difficulty of assessing performance in general, as various respondents complain. Furthermore, they miss an overall “body” to coordinate the performance target setting and reporting process. This, however, seems difficult to achieve due to the strong independence of the ministries.

However, one respondent did tell about a severe case of consequences due to deliberate cheating regarding the performance targets and appropriations needed therefore. The head of a large agency apparently forged the numbers to receive a substantially larger budget for his agency. Only after the draft budget had been made, the government detected this fraud, which obviously brought the ministry in a difficult situation. Consequently, the minister terminated the employment of the head of the agency. At the same time, the interviewee emphasised that the case was an isolated incident and had not occurred before.

However, respondents talked not only about (missing) consequences for under-achievements. They also appear to miss *“positive signals”* for good performance (R6). In fact, good performance (or even “over”-performance) might even be discouraged when funding is permanently reduced. It is a logical consequence that this circumstance affects how performance targets are set.

A respondent used the example of the funding for universities based on the number of students to demonstrate the difficulty of proposing appropriate indicators. In this instance, the universities receive funds according to the number of students. During the

year, the students' well-being and success appear to be somewhat neglected; the focus is on maintain a high number of students around the "census day" in autumn, according to the interviewee. He explains that the current system makes the education system, or more particularly, the schools and universities, follow the funding. This, however, does not provide any information about the impact or the quality of the education system as such. Consequently, one suggestion could be to count how many students find a job after finishing university. This, so the respondent, "*would have an impact for the education policy*". However, this target also has pitfalls:

"But then what the universities would create, they would create a private company which would hire the students from that university for a one year period, and then they would give them more funding because they would show nice results. So you have to always be careful when you create some kind of policy based on indicators, and performance indicators, so that actors are not creating any fuzzy, eh, how would I say, operations." (R3)

Another indicated factor is the transparency of performance information. Finland, as member of the Open Government Partnership, publishes a significant amount of performance information online. In addition, the information system Netra is accessible by everyone (with Finnish language skills). This is supposed to increase accountability towards politicians (together with the performance reports) and towards the public. It is argued that taxpayers have a right to know whether their money is effectively and efficiently used. Nevertheless, this openness could also bring an unintended outcome. A respondent suspects that "*maybe that's the reason why the information is so, kind of, in many cases quite dull, or, you know, kind of, it doesn't tell you much.*" (R2) Clearly, there are two sides to every coin.

As such, several respondents propose that maybe it would be time for a change in the system and the general underlying "doctrine" of performance management in Finland. However, as has been shown earlier, the organisational environment – structures and cultures – play an important role, together with a range of different interests of different stakeholders and government levels.

Regarding the hard use of performance information, only one respondent explicitly refers to corrective measures taken after negative performance results, such as cuts in appropriations or money, and – in exceptional cases – changes in management. However, also this respondent holds the view that in general it is a rather soft approach.

What does seem to have an impact, though, are the reports of the National Audit Office and the Auditing Committee. The assessments are taken and followed up seriously. A significant amount of performance information is also entered into the State Performance Information System Netra. However, the actual impact of the information provided there appears to be questionable. Possible reasons for this are touched upon earlier, which may include the low quality of the indicators, the missing link between relevant indicators and objectives, and the focus on quantitative, measurable indicators, at the expense of relevant ones.

To better assess the performance of the different ministries and agencies, and to make them more comparable with each other, more ministries and agencies are introducing a “traffic light system” to assess the results of performance targets. Generally, the opinion of such a system is quite positive. Respondents expect the system to become more transparent and more uniform, as one interviewee explains:

“it’s a good step forward if we have one scale to be used everywhere, and then we can start the discussions, why in the hell everybody has green lights, are you really so good, or should you give more difficult targets for yourself, or are you doing the assessment right”. (R4)

However, this quote also demonstrates the difficulties of such a system. For instance, according to which criteria are targets and lights assessed? Moreover, can one measure the relevant targets according to a three-colour system? A respondent is sceptical. In the case of the unemployment programme outlined earlier,

“That would be green. Green light in the traffic light system. You have the three months rule, you count the youngsters that go through, and you met that kind of criterion. It’s a green light. But the quality of the operations is, how would I say, mmh, it’s not reported. The traffic light system would be ok, if the indicator would be different.” (R3)

As such, while it is likely that this system increases the transparency of the objectives, it does not change the general, underlying problem with the objectives. It addresses the symptoms, but not the roots of the issues at stake. If, however, indicators of the quality of the actions were introduced, it would look differently, according to several respondents. Furthermore, according to some interviewees, it might not be successful if the budgetary and funding process and the performance indicators and objectives are not linked.

In brief, it is in performance agreements that performance targets and indicators are set. This is the case in all ministries. However, the quality and usefulness of these indicators has been questioned across the board, due to several reasons: low connection between indicator and target to be measured, focus on input and output instead of outcome, no real or “encouraging” consequences resulting from the analysis of the performance information, and finally, simply a lack of (perceived) interest in such information. The latter can be frustrating, considering the effort and time invested into formulating the performance agreements. However, just because the performance reports do not receive the (desired) attention in, for example, parliament, this does not necessarily mean that the information is ignored. A respondent said encouragingly that there are always various levels (both visible and invisible to the outside) where information can be used.

6.1.5 On the structure and culture of the performance management system

The previous sections analysed the perception of the purpose and the design of the performance management system. The use of performance information and the performance reporting are also illustrated. These sections show that the system consists

of various dimensions, and different stakeholders have different interests. Some interesting dynamics emerge, which are examined more closely in the present subsection. For example, the interviewees seem to have their own or shared ideas about the current shortcomings of the system and that something would have to change in order for the system to live up to its expectations. They also see what it would take to achieve that kind of system. These points are addressed in the following paragraphs.

First, a point raised by various respondents is the perceived need for a transformation of the mind-set within public administration regarding the performance management. According to the respondents representing that view, this mind shift needs to occur on several levels.

Regarding the indicators, proponents argue in favour of a complete revision of the indicator system. One needs to evaluate carefully which indicators work well, which indicators measure what they are supposed to measure, and, most importantly, the impact or effect of these indicators. Several examples are mentioned that make the usefulness of certain indicators questionable, such as the employment programme or the university funding.

Another crucial issue mentioned is the character of the system as a whole. Generally, a distinction can be made between systems based on trust and systems based on control. Control-based systems, as the one currently in place, rely on performance agreements and reports to assess whether organisations fulfil their tasks. In such a system, indicators and performance contracts play an indispensable role. According to a respondent, control-based systems are considered to be based on mistrust:

“Within the OECD and everything we talk about trust, you know, how to build trust, because all these bureaucratic kind of control systems they are, of course the very basic idea is of course, based on mistrust between different actors, and I think that that’s why you have to have rules, you have to have, you know, very kind of, nowadays, technical settings for these systems.” (R2)

However, at the same time there is a trend towards more trust-based systems, as discussed further in the following subsection. Trust-based systems are based on dialogue and interaction of the members of the organisation(s). Indicators and performance agreements play a role insofar as they provide some orientation where the journey is heading. However, performance results are not primarily used to hold organisations accountable with the accompanying sanctions or rewards. Rather, they are used to make the organisation grow internally, to learn from experience, and to constantly improve the organisation’s performance.

Among the interviewees, there were proponents of a stricter sanctions and reward system, and proponents of a trust-based system built on dialogue and interaction among employees. One interviewee questioned the label “performance management system”, if performance is not linked to any kind of reward system, or “positive signals for good performers”. What would the incentive then be to perform well? Another interviewee strongly emphasised issues of trust and true leadership, what he calls the cultural dimension of the system. Without trust and true leadership, an organisation is unlikely

to progress, according to his point of view. He claims that heads of units and ministries should be more courageous and learn to trust their employees. Apparently, certain pilot projects had not commenced, because from the ministry side there were constraints about the potential success and usefulness of that pilot. This, however, contradicts the idea of a “pilot project”. Another instance that leaves room for improvement is the performance agreement negotiations. According to an interviewee, these negotiations would need much more interaction, true dialogue, and exchange of ideas. They could be seen as a “process”, rather than a one-time event, where all issues can be put on the table. Apparently, agencies ask more policy-relevant issues to be discussed in these meetings and to be included in the performance agreements as well, whereas the ministries appear to be hesitant to these ideas.

In summary, an increasing number of organisations see the need for a changed system. However, in order to change the system, the mind-set needs to be changed. As Senge (2006, p. 12) states: “At the heart of a learning organisation is a shift of mind – from seeing ourselves as separate from the world to connected to the world, from seeing problems as caused by someone or something ‘out there’ to seeing how our own actions create the problems we experience.” Thus, there is no “out there” – it all starts within; within the organisation, with the individuals running and acting within the organisation. Therefore, if an organisation wants to proceed to more of a learning system, it has to create the basic conditions to achieve such a step. The mind-set has to change, and people have to be open to such a change. Moreover, trust plays a vital role. Trust needs to be placed in other employees, but there is also a need for the courage to relinquish certain power, to genuinely work together, rather than “against each other”. A weakening of the importance of hierarchy and official power attributions do not automatically imply a lack of respect or recognition. In a trusting environment, hierarchy and power attributions do not play any role. People trust and respect each other for what they do, rather than for the position they have in the organisation and the salary at the end of the month.

6.1.6 On learning forums

Learning forums are a major characteristic of learning systems. In the Finnish case, the respondents explain that certain networks or “learning forums” facilitate the exchange of ideas, allow for dialogue, and enable interaction among colleagues. The different learning forums are now examined further.

Firstly, there is a Performance Management Network. This network theoretically consists of 150 persons, 100-120 of which usually take part in one meeting. The participants are representatives of the ministries in charge of public administration, the “practitioners” mostly, and representatives of the operational or management level of the agencies. These meetings take place every six months and last about four hours.

The representatives of a ministry participating in these meetings describe the atmosphere in this network as “non-confrontational”. According to them, this may have to do with the fact that operative questions are usually not addressed during these meetings, and general issues are discussed. They include, for example, questions about

the development of performance management and the exchange of best practices and experiences, such as pilot projects. Participants consider it useful, though, to learn from each other's experiences.

Secondly, there is a "working committee", which is – according to an interviewee participating in it – "*a bit more formal*" (R4). In this working committee, there is one representative for each ministry, which means 12 people in total. They meet up once or twice every six months. The participants are from the "development side" of the ministries. They know each other well and consequently have open discussions: "*There is no reason to be silent in those forums*", as the respondent (R4) claims. However, during these meetings, there is much discussion, but the solutions are usually addressed outside this forum. The Ministry of Finance facilitates the solution finding.

Thirdly, there is the performance management training programme, led by HAUS, the Finnish Institute for Public Management. This programme takes place every six months. Ministries and agencies take part in these meetings, with the representatives coming from the operational level.

In addition, there is the secretariat for the development of performance management, which includes the following: Representatives of the Ministry of Finance, the State Treasury, and HAUS form part of this network. They meet at least once a month to discuss current issues and to facilitate the next steps in the ongoing performance management reform.

Furthermore, there is a network of the Permanent Secretaries, chaired by the State Secretary of the Prime Minister, and a network of "strategy people" from different ministries.

The former has developed from a "*discussion club*" to a "*real management group of the government*" over the years (R2). According to several respondents, this is mainly due to the personal effort of the State Secretary of the Prime Minister, who invited outside consultants to help build a team atmosphere in this group. When asked how he achieved that, and what his motives and evaluation were, he wrote, "there is a clear need for a more holistic approach for the government and the ministries working together". According to him, the existing organisational structures are only compatible with such an approach, if three things are provided:

"1) Forum for cooperation, integrating the different policies and creating ... a common view, 2) processes and culture that support a holistic and strategic approach, and 3) trust among the different actors and individuals."

Consequently, he put the creation of trust at the centre of his efforts. He engaged himself with the team and started a joint personal development project with the aim to increase each other's abilities, "*as individual leaders, as leaders of our own leadership teams, and as members of our group*". According to him, "*doing and learning together is a good way to create trust.*" This atmosphere of trust enables the team to discuss processes, propose new ideas about the organisation, and develop a joint agenda.

According to the State Secretary, this approach has been quite successful. There appears to be a “*big change*” in the group. Interestingly, “*everybody has given up his/her own personnel and decision power in order that we will work better together*”. This is an illustrative example of the role individuals play in the whole system. This person saw the goals and the potential of the organisation more clearly, and thus managed to present and live his personal vision in a way, which resonated with the other members of the team (in that case, the network of Permanent Secretaries).

The other network already alluded to is the network of the “strategy people” of the different ministries. This network is informal, with no formal powers. In this network, everything related to coordination matters is discussed. In general, both the network of the Permanent Secretaries and the one of the “strategic people” are considered quite informal. The open atmosphere allows people to discuss critically.

Apart from these cross-sectional meetings, every ministry and agency appears to have its own, internal networks as well, where targets, processes, and other relevant issues are discussed. Generally, all respondents highlight the non-confrontational atmosphere during these meetings, which appears to be because participants of these meetings have known each other for a while, and thus, trust each other.

The remaining question, however, is to what extent the results of these meetings are directly implemented. From the interviews, it appears that there is room for extensive discussion and dialogue, but the final decisions seem usually to be taken outside these forums at the top level of the organisation. Moreover, often the issues talked about remain general, and as such, probably do not provoke too much contestation.

Another learning initiative worth mentioning is the performance management training programme organised by HAUS. It consists of a three days’ programme, with two months’ time after the first training day, during which participants work on their own project ideas. A coach from HAUS supports them during this process. There are about 20-50 participants per training cycle. When participants meet after two months, they exchange best practices. The idea behind this approach is to learn from each other, and change attitudes smoothly. According to one representative of HAUS, it is important to “*get everyone on board*”. The participants are experts in their respective fields, which does not necessarily mean that they hold senior positions within their organisations. An important task of these participants is then also to “*change attitude inside their organisations*”, which is considered indispensable, if the administration wants to move forward.

Currently, the topics addressed during these training workshops are related to the reform of the performance management system. Agencies and ministries are encouraged to develop pilot projects concerning their own needs in the scope of the four main aims of the performance management reform: how to make the system more strategic, more horizontal, more uniform, and more integrated. Thus, in this training programme learning occurs primarily via the exchange of best practices, presentation of pilot projects, and the discussions about those issues.

The general philosophy behind the training programme fits into the learning and dialogue approach. Rather than giving an order top-down on how to implement and run the new system, HAUS and the Ministry of Finance count on a “*smooth transition*” and “*learning by each other*”. In so doing, they hope to change the attitudes of the civil servants, as logically, not everyone would be open to changes, who will then commit themselves to the changes and new approaches.

To summarise, there appears to be substantial support for horizontal and cross-sectional working. Literally all respondents interviewed were involved in one or more of these networking, or, learning, forums. However, at the same time, it appears that the most important decisions are taken outside this forum. As such, there would still be some room for improvement.

6.2. Finland – torn between accountability and learning?

The previous subsections have shown the complexity and multidimensionality of the Finnish performance management system. There are certain dimensions on which all or the large majority of the respondents agree, such as the need to adjust the system to contemporary challenges, making it more flexible and reducing the administrative burden. However, the opinions of the respondents diverge on other dimensions. This includes, for example, the use of performance information.

Generally, performance information is perceived to be used in a primarily soft way – notable exceptions of “extreme cases” excluded. This would speak in favour of a learning system. Room for discussion, dialogue, and thorough analysis is provided before any decisions based on the information are taken. In addition, respondents claim that it would be erroneous to describe the performance management system in Finland as an accountability system. Nevertheless, performance contracts – *the schoolbook* example of an accountability system – form the core of the performance management system. Does this not sound contradictory?

Indeed, the analysis of the Finnish performance management system has shown elements of both accountability systems and learning systems. While the general structure and underlying assumptions are largely still based on the “traditional” model of accountability, a shift of mind seems to have started taking place. Notably, also “officially” the Prime Minister’s Office and the Ministry of Finance explicitly promote a change towards “understanding systems” based on learning, reflection, and experiments, as stated in the Governments of the Future Project (Prime Minister’s Office & Ministry of Finance, 2013). Also in practice, the importance of dialogue, interaction, and cross-sectional cooperation is gaining greater importance. In every ministry and agency there are certain learning forums in place, where information is exchanged, members are considered to be in an equal position, and dialogue takes place – hence, learning forums in the sense of Moynihan's (2008) conceptualisation, and Senge's (2006) “discipline” of dialogue and team learning. However, there is still room for improvement. As mentioned by several interviewees, still more decision-making power could be handed over to these

dialogue forums. Especially the team leaders and heads of unit should not be afraid of losing power, but encourage their employees to think critically, discuss openly, and propose creative and innovative solutions to pending problems in the organisation and beyond.

Regarding the elements of learning organisations described in section 3.3.3, to greater or lesser extent all elements appear to be present in the Finnish performance management system. Learning forums and dialogue are described above. As concerns a shared vision regarding the future of the performance management system, there are two different approaches. On the one hand, we find proponents of a more top-down, cascading down, accountability system, where performance management is based on quantitative indicators. On the other hand, we also find proponents of a more horizontal system, where performance indicators should be critically assessed, the system even “*broken down completely*”, and be replaced by a system of qualitative, relevant indicators, where performance management is based on exchange of best practices, pilot projects, and dialogue across ministries, agencies, and units. Next, “mental models” describe our assumptions and ideas about how we see the world and how we act in it. They tell us much about how we work within and run an organisation. The constraint to start certain pilot projects, as briefly mentioned earlier, also refers to a certain kind of mental model. As various interviewees related, the fear of letting go, giving up control, letting the employees think independently and out of the box, is still too strong in certain organisations at certain management levels for real change to occur. Furthermore, as shown in the example of the Secretary of the Prime Minister or the stories of respondents who call for a “*substantial rethinking of the system*”, often it is individual efforts, rather than structural changes, that move an organisation forward. This is what would qualify as “personal mastery” in Senge’s (2006) terms. Finally, “systems thinking”, as the overarching discipline of the learning organisation, is less easy to detect. It can, however, be recognised in the overall, encompassing attempts to change the performance management system, establish shared goals, and recognise the contribution of each of the units, departments, agencies, and ministries to the success of reaching the overall goals.

As concerns the use of performance information, I find evidence for the four main uses Moynihan describes i.e. performance information is used politically, as only what is publicly accepted appears to be reported. Furthermore, it seems to promote political ideas. The perverse use of performance information becomes obvious in the stories about the indicators for universities and the unemployment programme, which indicate negative impacts. Importantly, these perverse effects occur in cases where performance results are tightly coupled to consequences, in that case, funding. Passive use of performance information can be found in the stories of performance information being collected in the scope of the performance management system, but not being used. Information appears to be collected and reported, but afterwards no further events follow. In this case, it seems that performance information does not have a considerable impact on the operations of the organisation. Then again, other respondents believe that performance information is indeed used internally, but not externally or for

accountability purposes. Hence, there appears to be a tension between the internal and the external use. When performance information is used internally, it is either used passively (with no direct consequences) or for internal improvement purposes (recall the improvement of the sick leave rates in one organisation). Externally, however, performance information can lead to perverse effects, in particular when tightly coupled to judgement or other consequences. This brings us back to the initial assumption that indicator-based accountability and learning do not fit well together.

Moreover, it is acknowledged that there is an urge to move towards more purposeful use of information. Information should be used for designing the operations and policy-planning. To best use the information collected, it might thus be advisable to reconsider the current approach towards performance management in general, and performance information in particular, and to engage more in double-loop learning activities. What is its real purpose? Is the information collected merely to fit into the standards of an indicator-based performance management system, focused on accountability? Alternatively, would there be an option to redesign the system, to consider alternative types of performance information, to dare something new? Would it make sense to reconsider also common views on accountability, and think of alternative ways of accountability? To reflect on these potentially considerable changes of the system involvement at all levels will be one of the critical factors of success.

6.3. Use of performance information in Lithuania

6.3.1 On the purpose of the performance management system

As mentioned earlier, the Lithuanian administrative culture is characterised by legalistic and hierarchical structures. This fact is also reflected in the perception of the performance management system. When respondents were asked about the purpose of the performance management system according to their points of view, the answers remained vague. Some interviewees see the purpose mostly in the alignment of resources and priorities, others mention the helpfulness of targets, goals, and indicators for defining and controlling the direction and success of budget programmes.

However, several respondents were surprised by this question and did not have a direct answer available. They referred me to the Office of the Government instead as the institution that set up the system and would know more about it: *“This is not a question for us, this is a question for the Office of the Government”* (R1). The respective respondents’ organisations merely put the information in the system, which would then be used by the Office of the Government for the reporting.

On the other hand, a respondent raised the question of why everything has become so complex, while things could be so much simpler:

“[The] strategic planning system is not something like NASA, where everything is so difficult. Just you have to understand what you want to have, where you are going, what you have to do, and what you have to reach, and that’s it.” (R2)

To make the system more user-friendly and clearer, several changes have been proposed and implemented during various reforms. In particular, the strategic planning system has been aligned closely to the budget process, and there are a detailed methodology and guidelines available to help with setting the indicators and targets. As a respondent claims, this made the strategic planning documents “more realistic”.

Nevertheless, at the same time, this “prescribed” form of performance management also makes the system rigid. A respondent describes the pitfalls of such a rigid system:

“In our country it’s mandatory to use it [the system], and in some cases it’s, I would say, with regard, everything is prescribed, the form of it, what should be the hierarchy of the, let’s say, which is the structure of the strategic plan, which sometimes ... even doesn’t make sense, because for example, Ministry of Finance, it has to give, let’s say, EU budget, some big amount of money, and it’s in separate programme, and you have to set the indicators for the programme and so on, but it doesn’t make sense, because you have to transfer certain amounts to the EU budget, that’s all. It is calculated on some formular, you calculate it and you transfer it. What are the results of this programme? But you have to make them. So sometimes it becomes a bit artificial because you have to do it, but on the other hand you see that it doesn’t make sense. So it is maybe good some kind of structure and similar across the public sector, but when it’s not so flexible, again, you get some kind of drawbacks.” (R3)

In general, the way respondents describe the purpose of the performance management system reflects a perception of the performance management system as being something organised at the top, with no or little involvement of those who are to use and implement it. Consequently, the system is perceived as burdensome and the true purpose is not clear to everyone. On the other hand, an outside source claims that this could also have to do something with the culture within the organisations and among employees; people do what they are told to do in an almost “military style”, without questioning the reason. I return to the cultural elements in section 6.3.5.

6.3.2 On the design of the performance management system

The functioning of the system is laid down in the Strategic Planning Methodology, a central document which is referred to by all respondents (Government of the Republic of Lithuania, 2002). The Strategic Planning Methodology suggests in a detailed way the type of performance indicators each organisation is supposed to have, and criteria how to develop (and assess) these kind of indicators. However, according to the respondents, this makes it a rigid and inflexible system. Furthermore, respondents complain about the overload of indicators and reporting obligations – to the extent that they feel they have to neglect their “real tasks” to comply with the rules set in the Strategic Planning Methodology. The following interview with two respondents illustrates this point:

“R4: Too much. I think that we have too many strategy documents. ...

R5: Yes, I don’t know if there is such practice in other countries, but it seems to me that in our department we have a lot of work, important work to do, and very

important initiatives that we have to implement. But a lot of time we are spending for the planning, reporting, and...

R4: filling forms, and so on.

R5: It's very big burden for us, and of course it doesn't help to reach the real results that we are responsible for. So we really pay a lot of attention to the planning and reporting, because Office of the Government required to provide information on time, and it's strongly controlled, but it's really, I think it's not efficient."

Consequently, several respondents perceive the control executed by their supervisory institution (in particular, the Office of the Government) to be strong.

However, not all of the respondents share this criticism. These respondents argue that there is indeed an overflow of indicators and lengthy reports, but this is not something caused directly by the guidelines in the planning methodology, as a respondent argues:

"[The] strategic planning methodology is not that rigid that it would require the ministry to come up with 100s of indicators. That's their choice. Basically it reflects their internal weakness. If the ministry as a whole is coherent, and the Minister has a vision, then you will have basically have a UK type, with a clear business plan of the ministry, few indicators, and the rest could be followed, I don't know, as a context or in the background. But if the ministry is a complex set-up, meaning there are several policy domains, and there is no strong leadership, then naturally bureaucracy fights for reflection of its own work in the strategic plan, thinking that if you are not in the strategic plan, then you will not get resources. And then these huge monster plans are basically a reflection of internal weaknesses. ... But emanating also from the lack of political leadership. So if the Minister does not formulate or clearly articulate and lead this process, naturally it's done by bureaucracy, which is fighting for its own place under the sun." (R6)

The lack of political leadership or interest is a factor mentioned in many interviews. At the same time, however, the strategic planning methodology is considered to have led to a narrow-minded focus within the organisations. Every goal and target is described in terms required by the methodology, while losing the bigger picture of what the purpose of the methodology was in the first place:

"I mean, forget the methodology. Can you describe to us or to me or to the Office of the Government what you want to achieve without [the] methodology? If you're able to describe and put goals and some sort of quantitative or qualitative [targets], if you want to steer budget and resources, that's strategic planning. And the rest can be worked from that. But if you are not able, you would see it as a burden. So the problem is not in the methodology." (R6)

To understand these contrasting point of views, I returned to the documents at stake, namely the Strategic Planning Methodology (Government of the Republic of Lithuania, 2002) and the "Methodology for formulating and applying performance measures used in strategic planning documents" (Office of the Prime Minister of The Republic of Lithuania, 2011). The methodology for formulating performance measures provides a

detailed plan on how to formulate and categorise performance measures in accordance with the government priorities and strategic goals. It does not require a specified number of indicators; however, it recommends to use an equal amount of quantitative and qualitative indicators. Nevertheless, depending on the specific situation, more or less quantitative or qualitative indicators might be required and considered more sensible. Forty percent of the indicators should reflect outcomes (i.e. strategic goals) and results (i.e. programme goals) against 60 % reflecting output (implementation of programme objectives) indicators (Office of the Prime Minister of The Republic of Lithuania, 2011). Furthermore, this guideline asks for a performance measure for every objective, programme, or strategic goal. However, the document specifically calls for a limitation in the number of indicators, while still providing comprehensive information. The Strategic Planning Methodology further specifies that on the output level, each objective should be measured by at least three performance indicators (Government of the Republic of Lithuania, 2002).

Thus, while detailed and extensive guidelines are available for how to set indicators and for which goals indicators are necessary, on paper these guidelines seem to leave enough room for each organisation to decide upon the number of goals and consequently, indicators. However, at the same time, something shared by all respondents is the difficulty of finding appropriate indicators.

“Of course there is a point [for having a performance management system]. Of course, just maybe there is like a problem, that in the governmental sector it’s kind of difficult [to find] the right criteria. The right criteria is very good thing. If for personnel, even for a person, for an institution as well, if you have criteria which really indicate whether your activity is going down or up, and it’s really useful for taking decisions, it’s ok. But sometimes we have a case with too many criteria. For example, there is, in this strategic planning methodology it says, ok, for measuring the goal, the strategic goal, you have to have at least one criterion. For measuring the task you have to have like three criteria. But maybe you don’t need three criteria for measuring tasks. So sometimes it’s just fulfilling the regulation ... But if you have a really useful, it’s another case.” (R7)

However, not only is it difficult to find the appropriate indicators, as various respondents explain. The actual use of performance information is not as straight forward as it might appear from the official strategic planning documents. This is examined in the following section.

6.3.3 On the use of performance information

A major challenge in Lithuania is that a significant amount of information is systematically collected, but it is recognised by academics, international organisations (such as the OECD, 2015) and the interviewees, that little use is made of this information. One respondent answers the question of whether the information put into the central information monitoring system (SIS) is used:

“Interviewer: And then who uses the information from the SIS system? Do you [your organisation] also use it, or...

R5: No, I just provide information, but I think this should be used by the Office of the Government, when they are reporting to the Government the results that are achieved by implementing different strategic documents, and I think that this system is created for the results-based management, and this information about the indicators and other information from the system should be used for the better solution of public governance improvement.

Interviewer: And do you have the feeling that it is used for this purpose?

R5: Personally, I, no.

Interviewer: ... And can you imagine maybe why it is not used?

R5: I don't know, maybe that we are far from this system, and we didn't feel the, we don't feel pulse of the system, how it's functioning in practice. Maybe, of course, we have a department in our [organisation] which is responsible for the strategic planning in our [organisation], for the preparation of the action plans of our [organisation], so this department deals and cooperates with the Office of the Government, the strategic planning division in this Office, with these colleagues, and of course, maybe they could better explain how the system is functioning. Because they are the users.”

Similarly, also another respondent is pessimistic about the use of the performance information:

“Interviewer: And would you say in general the information in the strategic planning, which is ... fed into the strategic planning system, is used?

R3: Good question. Good question. Sometimes, let's say, that it could be used better or more. So ... sometimes you could question is it something artificial that you just do because it is mandatory, or is it because you need it.”

These interview excerpts express what many other interviewees expressed. Calling to mind the discussion in the previous section on the perceived burden of the number of indicators and targets, many respondents complained that all the efforts concerning target-setting, collecting performance information, and eventual reporting, are of no benefit when they examine how these results are used, or rather, not used.

To a certain extent, however, this information is used for the ministries' reports to the government, and the government report to the parliament to inform about the progress and achievements of budgetary programmes. How far, however, this information is really “useful” and “telling”, is an aspect that has been questioned by several respondents. In fact, various interviewees felt that they had to collect the information because “they have to do it”, without knowing the actual purpose behind it. However, in particular the Office of the Government and the Ministry of Finance (i.e. the top level of the performance management system) stress the importance of this information for

keeping track of and controlling the progress of budgetary programmes and the implementation of government priorities.

Another important aspect concerning the use of performance information is the occurrence of negative effects. As in Finland, instances of “gaming and cheating”, such as data or indicator “manipulation” were reported by several interviewees. For example, respondents told about occasions where performance indicators were set in a less ambitious way to ensure their targets can be reached:

“But of course talking about the system in Lithuania, strategic planning system, I believe that, maybe, I don’t know, maybe 40 % of the data are safe. They are not lying, of course they are not lying, but it depends, you know. My colleagues are joking, there are lies, big lies, and statistics. So you can set one indicator, and feel safe all the time. But you decided to measure this indicator or to count this data not from the right point, and you will be safe all the time. So just, lies, big lies, statistics.” (R2)

As this respondent argues, choosing indicators wisely is supposed to be “safer”. Furthermore, it could avoid unnecessary, uncomfortable conversations with the superiors.

In line with this is also the deliberate selection of (positive) performance information for the performance reports, as public and political accountability is perceived to be quite strong. This issue is looked at in more detail in the following section.

Another interesting instance occurred in an organisation, where respondents did not seem to feel “safe” in telling me their experiences with the performance management system. In particular, they were careful in formulating potentially negative impressions. We were discussing whether organisations adjust their indicators when they see their organisation over- or underachieve them, when one of the respondent replied:

“Of course. But for the next year. Because, you know, in the system, we can’t change the values of the indicators during the year. Only if they are the product indicators, only in this case, so... We could make our result better, you know, if we had an option to change the annual value – it’s a joke, of course.” (R1)

While this respondent clearly emphasised that the remark was a joke, (and to me showing the “fear” of having been too open about something), at the same time it shows the potential vulnerability of such a system.

Furthermore, information is also instrumentalised for political purposes. The use of performance information is apparently highly dependent on the political will and interest of the respective politicians in power. While some politicians appear to value this information, and encourage the respective ministries to organise internal debates to discuss this information, others appear to have different priorities. However, as discussed in the previous section, when political leadership is disconnected from the performance management process, this supposedly leads to a lack of vision and purpose in the ministry concerned, and this in turn makes employees experience the performance management system as rigid, inflexible, and burdensome. On the other hand, some respondents also complain that while the information is available, politicians do not

necessarily use this information – they base their decisions on “emotional arguments” and only ask for performance information if it supports their argumentation and political motives:

“Political debates are [still] more powerful in this process, than data. Of course they use data, but not very deep, not always. Sometimes they would like only the, please find data to support our opinion, you know, it’s different. But it’s growing. If to compare situation, for example, 10 or 15 years ago, it’s completely different.” (R8)

Overall, it can be said that performance information is not used to its full potential. While there is a substantial demand from the top administrative level for this information, and all government organisations do their best to deliver this information, political demand still remains low. As such, the question remains, to which extent this information is useful and gives a complete portrait of the organisations at stake.

6.3.4 On the reporting of performance information

Regarding the reporting, the “Methodology for formulating performance indicators” suggests to report results on performance measures according to the needs of the account-holder. For example, the Government would be particularly interested in the implementation of the government priorities and other strategic goals. The Ministry of Finance puts more emphasis on effectiveness and budget-related results. In addition, the public and service consumers should be informed, in particular about information regarding service changes, improvements, attitude changes, or benefits to the public. Finally, the methodology also refers to the Seimas, European Union institutions and other external bodies that ask for certain monitoring data. Furthermore, in case of over- or underachievement of a specific aim, an explanatory note should be included in the semi-annual and annual performance reports (Office of the Prime Minister of The Republic of Lithuania, 2011).

Also in the opinion of the respondents, in particular due to the vast abundance of performance information, it is necessary to select this information according to the account-holders’ interests and needs, otherwise the main message would become lost in a sea of data. At the same time, I also heard a few times that it might not be too good to be open all the time. Certain data might be relevant for internal purposes, but would not communicate much to the outside world. Consequently, several organisations have established their own, internal performance management system – in addition to the national one. It might not come as a great surprise then, that they feel overburdened by the reporting requirements.

Nevertheless, an issue that emerged from most of the interviews, was the one of a perceived strong control, especially from the administrative forum, and to a certain extent also from the political forum: Accountability in the administrative forum, most explicitly expressed in the perceived control by the Office of the Government, appears to concern the interviewed employees. While the interviewees have not experienced occurrences of hard sanctions, such as budget cuts or programme closures, so far, they do experience “emotional punishment”. This includes feeling uncomfortable and guilty when

being held to account for sub-optimal performance or delayed achievements, as a respondent explains:

“Maybe sometimes our Chancellor, for example, if we are late very much, then not punishes, but they are very angry and of course it’s not very big pleasure to listen... So, it’s not punishment, but no, it’s not very big pleasure when you feel some guilt, how to say.” (R8)

The same concerns are expressed in relation to public accountability. Being portrayed in public reports as not achieving the promised results is something respondents (naturally) want to avoid.

However, here as well there are different versions of the story. A respondent explains that the link between reports and judgement is probably not as tight as perceived by the respondents. There would indeed be room for explaining why certain targets have not been reached sufficiently:

“[W]hen they say, they are reporting, then somehow they’re afraid they will be blamed for underachieving, even though it’s not their fault. They are still not perceiving it, but it’s really their role to analyse and to actually put proposals what to do, if they are underachieving...” (R6)

Next to the administrative and the political forum, accountability to the public plays a role in the thinking and acting of the Lithuanian administration, too. Also against the background of its membership of the Open Government Partnership, Lithuania publishes main (performance) reports online. The exception is the report on the efficiency of the back-office functions under the responsibility of the Ministry of Finance, as this concerns the internal management only. However, reactions from civil society, NGOs, or media generally remain scarce. Nevertheless, publishing these reports has provoked a certain hesitance among the respondents, who question its usefulness:

“Sometimes you know that with, especially with the politicians, they do not, not all of them, they discuss with you in a rational way. Sometimes they want to know, in an emotional way they want to have the discussion, or sometimes they do not go very deeply, and probably you will send the data, you will publish, but nobody will ask, why it’s so, or why it’s so good or so bad, and [the politicians] just, I don’t know, just will say, ok, you didn’t reach [the target] and you are bad. But you will not stay on the hill and do not shout why we didn’t reach, and why it’s not so bad. Because nobody wants to listen.” (R2)

This remark shows that there is a gap between accountability in the political and the administrative forum and their respective needs. While it is hoped that in the administrative forum there is more room for explanations and nuanced reporting, this is perceived to be less so in the political forum. This raises the question whether using the same reporting format for administrative and political accountability purposes is the most adequate form of accountability.

However, not only should be considered what forums to address, but also how. Simply making all information available is unlikely to lead to increased use. This is also claimed by the following respondent:

“Sometimes if you publish too many data, all data are like, become equal, and [...] for publishing you have to choose the main data. And not all of them should be published. Because if you have too many, they [the account-holders] will not see any data, they will be not interested in any data. [...] So sometimes, we believe and I believe, you need to have a package of your main data, what is important, because the address or the subject who is interested in these data, is very important.” (R2)

Thus, some respondents favour a more flexible approach towards the publication and the content of the data. Another interviewee also explains that publishing everything can even do more harm than good:

“Why to publish publicly how bad are [deleted for anonymity purposes]. We know this, and we are discussing this in a more, how to say, intimate place with them, but not for public very much to say about our problem. But it’s, you are finding place and time when what to say.” (R8)

This “intimate place” is what we could define as “learning forums”. In fact, the contrast between external and internal use became apparent in the interviews. When information is used externally, it makes the people or institutions involved feel uncomfortable in case of sub-optimal performance; the perceived pressure is higher. Furthermore, the tendency to adjust performance indicators, or, as one interviewee put it, the tendency to manipulate data increases. Less ambitious targets are set to ensure they can be reached by at least 90 %; easily measurable indicators are chosen rather than indicators that provide reliable information about the quality of actions, to name but a few strategies of such “manipulation”.

In contrast, when information is used internally, it is seen primarily as an improvement tool and less as an instrument for sanctions and thus creates a safer space for discussing controversial issues. However, the extent to which this information is internally discussed also varies across the institutions interviewed. While one ministry explains that under the previous government, many incentives and room for internal discussions about performance information existed, under the current government this is less so the case, and consequently, these discussions are less encouraged.

The discussions with the interviewees on the issue of reporting show that there is a great variety regarding performance assessment methods. However, they also revealed that while substantial amounts of information is collected and reported, there appears to be a lack of thorough analysis of the information. Hence, the perception of a “burdensome” system is present, where great effort is put into compiling data and reports, while the returns are not clearly visible to all respondents. Naturally, there are exceptions and examples where information is used in a constructive way and serves to stimulate debates. In most cases, this appears to be based on cultural and structural factors, (political) leadership, or personal efforts. Once more, this shows the importance of

considering the cultural and structural context to complete the picture. This is addressed in the following subsection.

6.3.5 On the structure and culture of the performance management system

Contrary to Finland, in Lithuania the envisioned direction of the performance management system was less uniform. The respondents expressed contrasting views regarding the performance management system, and the use of performance information in particular.

For example, a respondent calls for a more “human” approach towards solving problems, based on published performance information:

“... our Minister would like even for very bad [omitted] to close, to fire the heads of [omitted], and so on. He’s very strong. We are more humanistic, we would like to improve [omitted], not close.” (R8)

On the other hand, another interviewee claims that if sanctions were more strictly applied and public accountability utilised more, this could have a positive influence on internal processes.

“Let’s say, if you don’t achieve some results, and you get, let’s say, some media attention, punishing you for this, so of course, naturally, it would have some consequences. And then you would have to do something to explain why it is not achieved, and to do something ... If you are not getting, let’s say, punished for not achieving something, so what happens? So either you yourself decide that it’s bad and you try to achieve it, and you use this information as ok, I didn’t achieve it, why didn’t I achieve it, and you try to correct it, but it’s for your inside motivation. So yes, you have to, let’s say, know this information that something was not achieved, and look for the reasons. But if it doesn’t get attention and you don’t have this motivation, does it have any consequences for your internal things? Probably not. So, then it becomes a bit of a formal process. So I think it’s maybe better.” (R3)

The contrast between these two positions could not be greater. On the one hand, advocates call for a more flexible approach towards the collection and reporting of performance information. According to these advocates, measuring and reporting on performance is perceived as an unnecessary burden, which prevents the respective organisations in the administration to commit themselves to the issues that “really matter”. Furthermore, the manipulation of performance information prevents uncomfortable questions about under-performance, resulting in unambitious targets for performance indicators and reporting primarily on positive results. Following this line of reasoning, a more flexible approach within the performance management system would thus contribute to more reliable, relevant, and meaningful performance information. Consequently, this would increase the use of this information, making the efforts of collecting and reporting on this information presumably more worthwhile and less burdensome.

On the other hand, there are opponents to such a flexible approach. A stricter system, with more accountability, particularly with sanctions for under-achievement and rewards for over-achievement, would contribute to a significant system instead of a “formalised process” without true meaning.

However, the (punishment) avoidance strategies in the current system may increase when sanctions and strict controls are implemented. In other words, an increase in control and accountability mechanisms is likely to lead to an increase in avoidance and manipulation strategies and dissatisfaction with the system.

Another issue expressed frequently is the lack of political interest in the performance information. While there could be greater demand from the political side, to the respondents it often seems that decisions are based on “emotions”, “values” or political interests rather than on factual evidence, as the following quotes from different organisations show:

“The use of evidence is, in a rational world, or a perfect, sorry, it’s not even an issue of rationality, in a perfect world, so decisions are based perhaps on evidence. But there are also values, there are also political considerations, and ... sometimes one can understand why politicians don’t need that evidence, or don’t want to hear. Because they have their own [ideas], I suppose, what should be done and how it should be done.” (R6)

“R9: You know, sometimes the decisions are based on political will. So not evidence-based (laughs). Yes, so for us it’s sometimes very difficult to agree with some things, but of course, you know, it is political decision, you should follow this decision. ...

R5: And, sometimes, political leaders are not asking us to provide evidence for the new decisions. ...

R9: Even if we have this evidence (everybody laughs) not to do what we are doing, but decisions are decisions.”

The lack of perceived political leadership is also linked to the lack of political interest. This factor may also be traced back to the cultural legacy of the communist regime. A respondent explains that the Office of the Government traditionally, since independence, but also still during Soviet time, had a secretarial function rather than being involved in policy-making. Policy proposals were “guided through the entire process”, without challenging them.

“[T]he culture so far was not there that the sectorial advisors would line up with their own opinions and then signal problems”,

Respondent (R6) argues:

“[T]his demand for challenge, based on facts, based on evidence, based on experiences from other countries, has to be created by the leadership, which is in the sense [the] Prime Minister, because they serve the Government, ... and the Secretary of the Office of Government, Chancellor as it is called in our system. If this demand is not there,

being a rationalist, why would you create trouble getting across ... between line ministry which expects it [the policy proposal] mostly to be adopted, the Prime Minister who sometimes is not asking for that.”

These cultural features might thus explain why performance information is not always used, or is perceived to be ignored. Furthermore, there appears to be a lack of thorough performance analyses. However, various efforts were undertaken to improve the system and to make it more results- and evidence-based. Furthermore, attitudes among employees and in the different organisations and units have changed. However, in this case, it seems that the prevailing culture – among some employees and in some units – makes it challenging for genuine change to take place. Nevertheless, there are people who favour changes and who propose initiatives. It might just entail allowing the “young and energetic” blood to circulate more freely among the institutions, suggests a respondent. Nevertheless, change will not happen overnight, as another respondent explains:

“I believe that it will change, because to understand to change your mind, or to change every mind, like minds of all state or all employees ... it takes some time.” (R2)

Another issue that, in the literature, is often related to the cultural legacy of the past (Nakrošis, 2001; Pivoras, 2008) is the lack of coordination among government institutions. Thus, as demonstrated in the following section, the evidence for “true” learning forums is scarce. In addition, on a general level, most respondents mention the silo structure of the ministries. In some cases, this goes even further and there appears to be little coordination among the different units within ministries. Consequently, this also leads to diverging opinions about inter-ministerial action plans, which are by definition meant to be horizontal, cross-sectorial action plans:

“So if, for example, our ministry or another ministry prepares such a document [inter-ministerial action plan], you have to make meetings, you know, invite colleagues from other ministries and to think what you know of, other indicators or measures together, you know? To achieve the goals of that document. So, but of course, mostly everybody separately thinks of their indicators. They think separately of the indicators, and the document is a common, you know, because it’s approved by the decision of the government. But it can be, but they have several meetings for the preparation of this document, you know, but it can be called like a common job, and common work of preparation of this document.” (R1)

Another respondent is critical about the usefulness of these inter-ministerial plans, as he feels that money from his ministry is spent on issues that are not its top priority. This hinders motivation to participate in an inter-ministerial action due to lack of incentive. Instead, participating ministries should receive additional money from the coordinating ministry:

“Our ministry or other ministries which are participating will be more motivated. It’s a system of incentives. I’m getting additional money. So I would like to participate. Now I am doing with my own money something for you! It’s crazy for incentives, for

motivation! Why?! I would like to use this money for my own priorities. It creates weird situation.” (R8)

It appears that there is still a long road to go towards more horizontal cooperation and coordination. How this reflects also in the existence of learning forums is discussed in the following subsection.

6.3.6 On learning forums

As mentioned, compared to Finland, there do not seem to be many “systematic”, routinized learning forums. In addition, cross-ministerial cooperation or meetings are less evident in Lithuania.

Nevertheless, intentions to create learning forums and foster cooperation and dialogue do exist. In the following, I examine illustrative examples provided by several respondents.

One interviewee contends that in his organisation they explicitly want to move away from a mentality of blaming and punishment towards a mentality of improving, giving advice, and consulting. His ministry has undergone an extensive transformation of its subordinate agencies. To engage in cooperation and exchange of experiences, his ministry established a so-called “committee of experts”, which consisted of the nine largest agencies, and has now been extended to 17. The idea was to discuss the progress of the reform, positive experiences or problems, or solutions they have found. The reform is already implemented, so these meetings are not as frequent as before. Nevertheless, participants still see the benefit of it, as the respondent tells:

“So, yeah, we’re getting kind of to the point where we have fewer and fewer of those meetings, and when we have them, sometimes it’s less discussion and more just, somebody presenting something new, but without a lot of discussion. But it’s still an opportunity to, you know, if something comes up, if there is a new question, then, you know, a lot of times they are very eager to have the opportunity to not just be ordered to do something, but to be able to participate and offer their own suggestions. ... sometimes they would bring up issues like that themselves, they say ‘ok we have this problem, I’m sure a lot of the other ... agencies have this problem, this is how we solved it, this is what, you know, the procedures that we introduced to be able to sort out which ones are high priorities, which ones aren’t’.” (R10)

While performance information and target achievement are presented in this forum, the respondent emphasises that this is never confrontational, but encourages the respective agencies to examine their procedures and their data to find explanations:

“We don’t really say, hey, you should be at this target, and ... you only had this target, you know. Instead we say, mmh, we’ve noticed this ... this weird thing, we’ve seen that the number of [omitted] is high, can you please explain why, can you think about why, could there be a problem with the data, you know, could there be some other issue going on. So, you know, it’s not really that confrontational, it’s more, you know, educational.” (R10)

Furthermore, the Office of the Government is in charge of several initiatives. The Office recently conducted a personal coaching programme for representatives from all ministries, particularly the State Secretaries, as they are formally responsible for the strategic planning processes. This coaching addressed issues such as evidence-based decision-making, use of performance information, and general strategic planning capacities.

Another important forum is the meeting of the strategic planning experts, which takes place on an informal, *ad hoc* basis. It is chaired by the Office of the Government (sometimes together with the Ministry of Finance), and strategic planning experts from all ministries participate. Asked about the atmosphere in this forum, a respondent replies:

“Interviewer: ... can people talk openly, critically, about these things?”

R3: Yes! Because it's kind of, quite, I would say, low level meetings, it's not like ministers are [present], or the vice ministers, so basically heads of unit, heads of departments, maximum, so basically they talk very openly and aggressively. So it's no problem about that. They say what they think.

Interviewer: ... aggressively, do you mean towards others, or...

R3: Eh, no, basically towards the system. Let's say, we don't like something, how it works, and yeah, we, usually, as I said, more initiative in the past years came from the Prime Minister's Office, and basically, usually they defend their position, and explain why we need it, and so on. Yeah, they try to cool it down and persuade that, why it is needed, and... And of course ministries they don't like changes, and more work, let's say. So they usually are opposed to that.”

As such, openness and criticism are freely voiced in an environment where the hierarchies are comparably low. However, this reaction also expresses the frustration with the system and the apparent lack of common vision of how the system could or ought to look in the future. At the same time, other respondents question the sincerity of the contributions of the participants, especially when discussing information from the political arena, and when voicing their concerns about distrust of participants at the political level. *“We think that it's a little bit gaming issues”*, says a respondent (R11).

Nonetheless, another respondent is positive about these meetings, in particular about the horizontal approach.

“Very useful is this opportunity to meet each other and to discuss problems, what we have in our ministries, to share experiences, to share solutions, how we are doing one or other things. For example, one ministry, the division like mine, they have very good relation with Minister, with politicians. We have not succeeded so much in this, so we shared some experience. But of course in the other ministry the head is a woman, and sometimes it's easier to go to the man, Minister, another relation than between man and man. So, it depends on many aspects, but in any case, you share. For example, in this case I use sometimes my deputy, who is a woman. ... So problems like these we are trying to solve, to change the situation, and then we discuss how to do. I like this.

It's really, it creates some culture of cooperation and open discussion, it's not like, centralised, you have to do this, no, we're trying to discuss and to find solutions together, so we, I like.” (R8)

The Office of the Government is planning to develop this strategic planning experts' forum to work towards a common vision and to strengthen horizontal, cross-sectorial cooperation. It has been noted that in focused meetings on specific issues there appears to be a greater demand (and supply) of performance information. A recent proposal concerns the establishment of an informal team of strategic planning experts (with the acronym SPEAK), which would focus on “trainings, administrative capacity-building, networking, informal best practice”. According to the respondent, this would help shift the focus from formal guidelines and methodology to the actual content and thus improve the efficiency and effectiveness of the performance management system. The aim is “to build a real community, a real network of strategic planning experts.”

In brief, learning forums, or networks, play a less prominent role in Lithuania than in Finland. This may be a logical consequence of the solidified silo structure of the ministries (and even departments or units within the ministries). Inter-ministerial cooperation is encouraged; however, the interviews indicate that this “cooperation” mainly exists on paper, i.e. in the strategic action plans.

Nevertheless, existing efforts to overcome this silo structure and to foster horizontal cooperation are highly appreciated by the participants. Moreover, the prevailing culture and structure appear to hinder these efforts to bear fruit. Interestingly, regarding the use of performance information in these forums, the more specific and concrete the discussion topics, the greater the demand (and supply) for performance information.

6.4. Lithuania – a clear case for accountability?

The analysis above demonstrates that, as expected, Lithuania has a meticulously planned performance management system in place, based on a strong hierarchy of performance targets and indicators, which are all fed into a centralised performance monitoring system. While most of the information is publicly available, feedback from the public sphere is scarce. Nevertheless, internal and political control is perceived as strong – to the extent that several incidents of manipulating data or results were reported during the interviews. To a certain extent, this contradicts findings from the literature, where it is claimed that in Lithuania political accountability is moderately low (Pivoras, 2008). However, at the same time, the interviewees acknowledge that no “real consequences” have been noticed in cases of under-performance. As such, *de facto* and perceived political accountability might simply be something different. Nevertheless, literature and the interviewees agree that this has an impact on the formulation of performance and policy targets (Pivoras, 2008).

Initiatives to encourage the use of performance information for learning do exist in some cases; however, cultural and structural features of the administration limit success. In particular, the communist legacy seems to play a role (Nakrošis, 2001; Pivoras, 2008,

2013), but Lithuania also benefits from EU Structural Funds, which requires extensive reporting and assessments.

Nevertheless, the general awareness of abundant performance information, but of low quality or not used, has led to several reform projects. The most recent effort, initiated by the Office of the Government, aims at encouraging managers to use available information. Furthermore, the establishment of a network of strategic planning experts is envisaged, which would encourage the exchange of experiences in a more personal setting.

If one considers the elements of learning systems, i.e. learning forums in which dialogue and team learning can take place, shared vision, mental models, systems thinking, and personal mastery, the results are somewhat negative. There are no “systematically” established forums for the exchange of experience and best practices, except for notable exceptions described in section 6.3.6. Apart from that, other networks also exist. However, they have not been outlined in the section on learning forums, as important elements described by Moynihan (2008) were not considered to be present. These include, most importantly, the “equal footing” of participants. In particular, in such a hierarchic and generally low-trust culture as in Lithuania, it is unlikely that participants will engage in genuine dialogue if superiors are present to whom participants are accountable. Furthermore, and in strong contrast to Finland, the Lithuanian respondents do not seem to have a clear vision about the system. This finding is in line with findings earlier in the scholarly literature (Pivoras, 2008).

However, regarding the mental models, the findings are intriguing. All respondents directly involved in the collecting and reporting of performance information shift blame to their direct or indirect superiors. There seems to be almost a “universal resistance” to this type of performance management system. On the other hand, respondents who are a little more distanced from, or at the top of that system, have a more balanced opinion of it. One has the impression that the people who implement the performance management system are not involved in its development, and rather receive top-down commands to implement it. That would also explain the lack of a shared vision or understanding of the purpose of the system.

Next, evidence of personal mastery examples is also scarce. While there are people who attempt to contribute to the system in a different way, there are no outstanding examples that have a clear impact on the running of the system. Nevertheless, other respondents recommended a ministry to me for its innovative approach to performance management. Interestingly, this ministry has only a limited number of indicators and goals in the strategic action plans, and operates a separate indicator system for its ministry. The respondent informed me that when performance management was introduced in Lithuania, his ministry was in an advantageous position as it already had a strategic plan dating to Soviet times. Therefore, working with a strategy was a familiar concept to this ministry, and it might be due to this that they managed to implement more easily the performance management system in their ministry.

Finally, the concept of systems thinking is not easily tangible. Nevertheless, there is generally – especially among the “developers” of the system – the desire to overcome the silo structures and to focus on horizontal, inter-ministerial cooperation. Accordingly, budget and strategic planning (or performance management) systems would be fully integrated and aligned according to the long-term goals of the country. However, results from the interviews indicate that much work remains. For the developers of the system, this would mean greater engagement with ministries and agencies, and the development of a simultaneously functioning system, rather than top-down, so that the real purpose of the system is clarified. For the implementers this would entail looking beyond their own ministry or agency, and seeing their contribution to the system and the functioning of the administration as part of the bigger picture.

Similar to Finland, in Lithuania there is also evidence of the four main uses of performance information as elaborated on by Moynihan. Performance is used politically, as politicians carefully select which information to use and which to ignore, in which cases to ask for supportive information, and in which not. Several respondents provided evidence for perverse use of performance information. Performance is used in a perverse way – as in Finland – especially in cases where accountability is perceived as strong. In these cases, performance is carefully selected for reporting, and “critical” information is mostly kept and dealt with internally. However, most performance information is used passively, according to respondents. A significant amount of information is collected and reported on, because law requires it. However, the effects of these performance reports or the use of information is generally less visible. Purposeful use is then also limited; nevertheless, the Office of the Government is investing great effort into remedying this situation.

In summary, we see that in Lithuania accountability is – as expected – the main purpose for the performance management system. This has to be understood in the context of Lithuania’s accession to the EU and NATO in 2004, which required making the budget reliable and keeping track of the performance (Nakrošis, 2008). However, this resulted in a steady increase in indicators to the extent that the supply of information exceeds its demand considerably. In turn, the usefulness, meaningfulness, and rigidity of the system are questioned on all fronts: the coordinators, the users, and the evaluators.

To solve this problem, considering the prevailing administrative culture is indispensable – the lack of trust and deference of authority is a challenge. Hence, any efforts to change the system and encourage more interaction are unlikely to succeed if these cultural characteristics are not taken into account. As one interviewee described it, perhaps the Office of the Government is ahead of its time regarding its recent efforts to move towards a more meaningful use of performance information. More circulation and “fresh blood” might be necessary to demonstrate that the road to learning is perhaps not as unpaved as they might think.

6.5. Conclusion and implications

Contrary to my expectations, accountability elements play an important role in the Finnish performance management system. Performance information is used largely for accountability purposes, with hierarchical reporting and the – even if not strong – possibility of being sanctioned for under-achievement. Performance indicators are mostly of quantitative nature to easily measure the fulfilment of targets. This, however, results in indicators measuring actions instead of quality, and indicators chosen based on their measurability rather than their relevance. The introduction of a traffic light system for performance indicators increases this effect; making performance information publicly available in a highly aggregated way creates incentives for “manipulating” indicators, such as lowering targets to ensure they can be reached easily. Several respondents refer to such incidents.

At the same time, great effort is invested in overcoming the lack of coordination among ministries and in fostering communication among employees. All interviewees report that their organisations are involved in one or another network. These networks are based on trust and a non-confrontational atmosphere. They facilitate the exchange of experiences and best practices, and are generally highly valued among interviewees. Nevertheless, evidence of performance information use in these networks is low.

Table 8 below summarises the comparison of the Finnish and the Lithuanian performance management systems.

As I perceived Finland as a most likely case, does this mean that if even Finland does not have a fully-fledged learning system in place, we should “abandon the idea of learning organisations” (Grieves, 2008) altogether? I suggest a more nuanced answer: If Finland does not manage to use performance indicators for accountability *and* learning purposes, we might conclude that it simply does not work like that. Performance indicators, especially when made public and reduced to quantitative information, are likely to counteract any learning efforts. Nevertheless, important learning efforts exist, such as the performance management training programme of HAUS. During this training, the emphasis is on the exchange of knowledge, experience, and best practices rather than on naming and shaming or other accountability mechanisms.

Thus, with reference to the propositions formulated in section 4.1, I see the initial assumption of the incompatibility between using performance information both for learning and for accountability confirmed. However, I do not claim that accountability is *per se* “bad”. On the contrary, accountability is necessary for a transparent and democratic society. Nevertheless, we might have to re-think our concepts of accountability and performance information.

Table 8 Comparison of findings for Finland and Lithuania

	Finland	Lithuania
Purpose of the system	Accountability, steering & control, guidance & feedback; but also criticism of NPM doctrine; form follows function	Diverse opinions, often purpose not clear for respondents
Design of the system	Top-down, but increasing efforts for horizontal cooperation; Performance contracts as core element; one-size-fits-all vs. tailor-made; trust as important element	Top-down, rigid system; Strategic Planning Methodology as central document
Use of performance information	Perverse, passive, political, and purposeful uses	Perverse, and in particular strong passive use, political use; purposeful use limited; difference between internal and external use
Reporting of performance information	“Soft” consequences, traffic light system; missing link between targets and reporting	Soft to hard consequences; perceived strong control, blame avoidance
Structure & culture of the system	Recognition of need for transformation of mind-set: complete revision of indicator system, from control-based to trust-based system	More flexible vs. more rigid system; lack of political interest and leadership; lack of (horizontal) cooperation; strong silo structure; lack of common vision; blame/responsibility shifting
Learning forums	Large variety of learning forums at all levels of administration: dialogue, interaction, trust, common projects	Rather sporadic; mistrust; recently, more efforts to cooperate and share experiences and knowledge
Type of learning	Single- and double-loop learning	Single-loop learning

The incompatibility becomes particularly obvious in the case of the Lithuanian performance management system. As expected, the Lithuanian system focuses mostly on accountability. The control from the account-holders is then also perceived as strong, to the extent that targets are set at lower levels than achievable, and preference is given to easily measurable (and achievable) targets. Elements of learning systems, as described in the conceptual chapter, are absent – with notable exceptions, of course. Contrary to the perception of strong control is the recognition that performance information is not

used beyond the reporting. However, to generalise this point would also not do justice to important initiatives and efforts to build analysing and assessment capacities that make better use of the information available. It seems that, instead, “safe” performance information is used for reporting, and hence, for accountability, while “critical” information is reserved and discussed “in more intimate places”, as a respondent explained. These intimate places could indeed be learning forums. However, they could also be one-to-one talks or informal conversations among employees. Importantly, it is expected that this information would not leave a certain “forum”. There seems to be a reluctance to expose this information to the public, or to political or higher administrative levels. Moreover, this is exactly the point: When certain, particularly “critical” performance information leaves a specific “safe” forum, organisations or individual employees make themselves vulnerable to critique. In particular, political and public accountability are unpredictable and, as such, organisations are deprived of the room for analysis and debate once the information is “out there”. Add to this a generally low-trust environment, and it is not surprising that respondents are sceptical about the usefulness and success of the performance management system.

In conclusion, my research has resulted in both expected and unexpected findings (see table 9 below). Performance information is used in both Lithuania and Finland primarily for accountability purposes. This is surprising as, based on the administrative culture, I would have expected more learning elements in the Finnish system. However, these are not completely absent in Finland. On the contrary, there are indeed many elements of learning systems present and operative, but learning and accountability processes are decoupled. Lithuania, on the other hand, hardly presents any established “learning forums”; the focus is clearly on accountability. However, recent efforts are attempting to reconcile this. How far these efforts will bear fruit remains to be seen. Based on the literature on learning, I contend that this is dependent on the degree to which participants are involved in the process and not simply presented with the “end product” without opportunity to see their opinions reflected. As one respondent remarked, in their organisation they do not feel “*the pulse of the system*” as it seems too removed.

Table 9 Propositions and findings

Propositions	Findings
Accountability and learning purposes are unlikely to be compatible in performance management systems.	☹
Finland, based on its administrative structure and culture, is a most likely case for learning elements in its performance management system.	☺
Lithuania, based on its administrative structure and culture, is a least likely case for learning elements in its performance management system.	☺

Finally yet importantly, I find that Lithuania engages predominantly in single-loop learning. Since the establishment of performance management systems in the country, efforts continuously seek to improve the existing system. However, the basic assumptions underlying this system have not been questioned adequately. The same applies to Finland, although with the increasing role of learning forums and informal exchanges, approaches towards more double-loop learning are apparent. Yet, in both Finland and Lithuania, the evidence of perverse effects resulting from indicator-based performance management systems has been shown. Perhaps it is time to rethink the performance management approach in its entirety, and attempt to find ways to do it differently.

7. CONCLUSION AND OUTLOOK

*“Two roads diverged in a wood, and I –
I took the one less travelled by,
And that has made all the difference.”
~ Robert Frost*

We have reached the final chapter of this dissertation. I first recap the main findings of this study and answers to the research questions, and discuss the theoretical contribution of this thesis. Then, the limitations and possible further research avenues are discussed. Regarding the practical implications, I address the four value-rational questions mentioned earlier, in line with the *phronetic* research approach: (1) Where are we going with performance management systems? (2) Who gains and who loses, by which mechanisms of power? (3) Is this development desirable? (4) What should we do?

7.1. Findings and theoretical contribution

This thesis departs from the observation that performance management, as undertaken following the NPM doctrine, has not lived up to its expectations. The literature on the negative effects and the empirical evidence support the assumption that in the presence of increased accountability and control, increased gaming and cheating occurs to avoid sanctions or negative reputation. At the same time, calls to use performance information to learn and to improve policies and management have become louder. However, empirical evidence for learning from performance information is scarce. This has led to the assumption that accountability and learning might simply be incompatible.

Against this background, **chapter 2** provides a review of the literature on the negative effects of performance management systems to determine why and under which circumstances these negative effects occur. I found that negative effects particularly occur when performance information is used in a hard way, i.e. when consequences and judgement are tightly coupled to performance results without sufficient room for interpretation and analysis. These negative effects consist of manipulating either outputs or measurement. Measurement manipulation refers to “statistics juggling”, which often does not directly affect the policy or the services provided. This includes, for example, indicator pollution, indicator inflation, misrepresentation, or misinterpretation. However, as the example of the scandal surrounding VW shows at the outset of this thesis, measurement manipulation can also have severe effects – particularly when combined with output manipulation. Indeed, output manipulation has more severe consequences as it affects service provision. These effects may include a tunnel vision (“teaching to the test”), myopia, measure fixation, and organisational paralysis. In the

empirical cases analysed in chapter 6, evidence for these negative effects has been found; measurement is manipulated by, for example, setting lower and more easily achievable performance targets, or by indicator pollution. Output is manipulated in the case of the unemployment programme, for instance. Rather than helping young people find a job, they are left on their own – yet the targets are officially reached. The same occurs in the case of university funding related to the number of students, and evidence for improvement of the university’s studying conditions seems to be scarce.

The theoretical contribution regarding the effects of performance information lies in the fact that I show that the complexity of the field, in particular the various forms dysfunctional effects can take, can be reduced to the question of “incentives”. The higher the incentives for good performance (e.g. receiving bonuses or higher budgets, but also avoiding blame and negative reporting), the more likely gaming and cheating are to occur. This is especially the case when performance information is used in a hard way, i.e. when it is tightly coupled to sanctions, judgements, or other consequences, either factual or perceived.

Consequently, as the manner in which performance information is used seems to play an important role, it seemed logical to examine more closely the literature available on this topic. This was accomplished in **chapter 3**. The literature on the topic, and analogically, the different ways in which performance information can be used, are abundant. Often the distinctions are nuances, however, which led me to “simplify” the categories by looking more in depth into two forms of use – accountability and learning. Both accountability and learning are known and widely studied concepts, and can be situated on opposite ends of a continuum. Accountability implies a hard use of performance information, and is more oriented towards controlling and explaining the past; learning, on the other hand, is more future-oriented and looks at how policies and management could be changed. Despite the abundant literature available, important shortcomings of this literature are identified. Only fairly recently did researchers begin to study (organisational) learning in the public sector. These contributions are mostly theoretical so far. While within the accountability literature there is a strand that suggests learning to be one form of accountability, literature on this topic is lacking, and empirical evidence even more so.

Consequently, I wanted to contribute to the link between the two strands of the literature both theoretically and empirically. I examined to what extent learning and/or accountability elements are present in two different performance management systems with different contexts, and the implications thereof. Furthermore, I provided insights into the theoretical discussion on accountability versus learning systems.

Empirically, I provided a qualitative perspective on the use of performance information in learning- and accountability-systems, thereby enriching the predominantly quantitative debate on this issue (e.g. Askim, Johnsen & Christophersen, 2008; Moynihan & Landuyt, 2009; Moynihan & Lavertu, 2012). The findings in **chapter 6** show that elements of accountability- and learning systems are present to varying degrees in both systems. However, important differences are detected.

Finland, the case I had conceived as a most likely case for learning, only partly confirmed my expectations. While many learning elements are present – not only *de jure*, but also *de facto* – accountability still plays an important role, notably in the performance contracts between ministries and agencies. This (perceived) accountability leads to data manipulations and strategic selection of indicators. In turn, this has an impact on the meaningfulness of these indicators. What do they really convey? Are they chosen to avoid criticism and uncomfortable situations of having to justify specific actions? Alternatively, do they present a realistic picture of the issues at stake?

In Lithuania, the least likely case for learning (and most likely for accountability), the findings confirmed my expectations more closely. Performance information is perceived to be used mainly for accountability purposes. Substantial performance information is collected and reported on, but no further actions result from this. Most respondents therefore question the usefulness and the meaningfulness of such a system. Furthermore, and in stark contrast to Finland, the purpose and meaning of the performance management system is unclear among the interviewees. While in Finland there appears to be a shared vision regarding the “ultimate goal” of the system, this is less so the case in Lithuania.

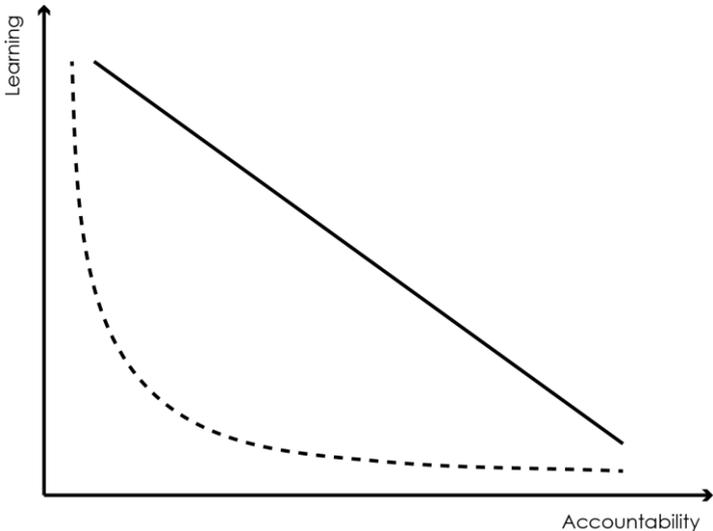
Interestingly, particularly in Lithuania, the cultural context seems to play an important role. Against the background of a communist legacy with low trust in bureaucracy and the political system and a strongly legalist culture, people appear to be more concerned with following the rules and fulfilling the legal requirements, rather than looking beyond and considering the bigger picture. Indeed, while in Finland the importance of the “bigger picture” was stressed several times, the Lithuanian respondents did not show strong signs of reflecting on this. The silo structure of the ministries and departments is then also more manifest in Lithuania than in Finland.

To summarise, this study has found that even in countries that I considered “most likely” cases for accountability respectively learning systems, elements of both these systems are present. This is not surprising, though, given the need for both accountability and learning in well-functioning administrations: Accountability in terms of answerability and responsibility towards stakeholders, politicians, and citizens; learning in terms of being better prepared for the future – because without learning, organisations are unlikely to improve.

There is little doubt about this general assumption. What is questioned, however, is the use of performance information in these systems. The case studies provide empirical support for the premise that using performance information for accountability is incompatible with simultaneously using it for learning purposes. This is mainly due to the different mechanisms at play. Once performance information is externally reported and judgement or other consequences are expected, there is no longer room for internal analysis and interpretation through dialogue and learning forums. This in turn leads to performance information being “adjusted” to the needs or expectations of the account-holders, and as such, in many cases the informational content is of little value.

Hence, this study also enriches the debate on the use of performance information in accountability- and learning systems theoretically. Figure 10 below demonstrates the described mechanism. We have seen that the more accountability prevails, the more dysfunctional effects are going to occur. This is usually thought of as a “linear” relationship between accountability and dysfunctional effects (the solid line in figure 10). However, the literature review and the empirical evidence suggest that this relationship might not be linear, but resembles a curve (dashed line in figure 10). This has important implications. Instead of a “the more learning, the less accountability” or “the more accountability, the less learning” relationship, it implies that even a minimal extent of accountability is already detrimental for learning efforts. As chapter 3 has shown, using performance information for learning or for accountability implies a substantially different mechanism. When performance information is used for learning purposes, it requires moments of dialogue (learning forums) in which performance information is discussed, analysed, and interpreted, before conclusions or judgements are made. Performance information is used in a soft way. In accountability systems, on the other hand, judgement is tightly coupled to the performance results. Hence, if performance information is supposed to be used for accountability purposes, it can no longer be used for learning purposes. This is related to the finding that the threshold for gaming and cheating to occur in accountability systems is low, and consequently, account-givers might already adjust their behaviour (or their data) to avoid blame or other negative consequences (see also Van Dooren et al., 2015).

Figure 10 Trade-off between accountability and learning



Source: Author's work.

Linking this back to the empirical findings, it might now be more comprehensible that even in Finland we still find a significant prevalence of accountability; and even in Lithuania we find some learning elements. In Finland, although much has been done to create a learning culture with different learning forums, pilot projects, and training programmes, apparently the accountability elements present in the system partly

“overrule” these learning efforts. In Lithuania, on the other hand, accountability clearly plays the predominant role. However, particularly as the performance management system in Lithuania has been implemented fairly recently, it is in a constant development phase, in which ways are to be found to make the system more meaningful and functional. Hence, learning to improve the system also plays a role in Lithuania, and it is thus only natural that learning efforts are undertaken. However, against the strong prevailing accountability culture, greater effort is needed to invest in a genuine learning culture to outweigh the negative effects such an accountability-based system bears with it.

This leads to another important aspect in this story, i.e. the implications for the design of performance measurement systems. As outlined in **chapter 3**, the use of performance information is strongly linked to the design of the system in which the information is used. Learning- and accountability-based systems thus have different requirements to be functional. While the former benefits from an approach that includes as many levels within the organisation (and other stakeholders) in the design of the system as possible, the latter might be more functional when the design and implementation are directed from the top. Moreover, in learning systems, dialogue and interaction are key to make use of the performance information provided and collected. This, in turn, requires a culture that facilitates such dialogue and interaction, and provides for blame-free space. Hence, trust, the feeling to be part of a team, a shared vision of the purpose of the system, and mutually shared values play an important role here. This, however, also implies that there is no “one size fits all” recipe available for the design and implementation of learning systems. Nevertheless, some crucial elements can be distilled, which have also been investigated, both theoretically and empirically, in this thesis. Consequently, if performance information is to be meaningful and useful, the users should move away from the “one size fits all” logic, and recognise the different mechanisms underlying different purposes of use. Here, the context, both structural and cultural, plays a decisive role.

Finally yet importantly, this thesis has also claimed that accountability is not *per se* a bad thing. On the contrary, accountability is a necessary element in any democratic system. It is thus rather a question of “what” kind of accountability, than of “whether or not”. When accountability is decoupled from performance indicators and organisations are held accountable for their learning efforts instead, this could be a courageous double-loop learning strategy. The literature on accountability for learning (Olejniczak & Newcomer, 2014; Van Dooren & Willems, forthcoming) as analysed in section 3.4 provides inspiration for such a route off the beaten track.

7.2. Limitations of this study and future research ideas

This study contributes both theoretically and empirically to the existing research on performance management, and in particular to the use of performance information. As in any scientific study, the scope of a PhD dissertation is limited by certain choices. I

explain these choices in the following paragraphs and suggest additional routes for research.

First, regarding the methodology, conducting case studies always entails the risk of receiving criticism concerning the replicability and generalisability of the study. I address both issues in turn.

In chapter 4, I explain that it is not intended that the case studies I conducted be replicable. As the findings of the case studies show, political interests and priorities seem to play an important role in the design, implementation, and use of performance management systems. Consequently, with any change in government (be it the governing party or simply a change in politicians), the results might also look different. Moreover, in April 2015 a new government came into power in Finland. It could thus be that some of the findings would already have to be reconsidered. Nevertheless, it has never been the aim of this research to provide “context- and time-independent truth”. Rather, I have explicitly chosen to conduct a comparative case study able to take the context into account, and offer “context-dependent, practical knowledge” instead (Flyvbjerg, 2001).

This links to another issue, often raised in the context of case studies, i.e. can one generalise from the findings of a case study? The answer to this question depends on the selection of the cases. In this thesis, I have chosen two cases that I would place at opposite ends of a continuum in order to draw some broader conclusions – a most likely and a least likely case for learning. Bearing the context-dependence in mind, I recall the “power of example” to suggest that there are indeed certain mechanisms which appear to influence the use of performance information for accountability or for learning purposes. The results indicate that the use of performance information is strongly connected to the organisational context and culture. Trust and political interests, but also the extent to which governments feel the “external” or international pressure to have a certain system in place, play an important role in the use of performance information. These findings are in line with scholarly contributions applying quantitative methods, most notably based on surveys (e.g. Askim, Johnsen & Christophersen, 2008; Moynihan & Landuyt, 2009; Moynihan & Lavertu, 2012). As such, in a high-trust, egalitarian, and cooperative environment where people share a vision and see themselves as contributor to a “bigger picture”, it is more likely that performance information is used for learning, than in the opposite case, where a low-trust, hierarchic and legalist environment is inductive for a dominant focus on accountability. The findings furthermore suggest that these two purposes of use, accountability and learning, are difficult, if not impossible, to combine.

Next, a few words about the case selection. A case study offers the opportunity to analyse a case in depth and move beyond the “official” information, digging deeper below the surface. This was undertaken in the two case studies on the Finnish and the Lithuanian public administrations. Nevertheless, my results show that there is also considerable diversity within the cases. Thus, while I can make claims about the “general” trend in the respective cases, they have to be restricted by the respective contexts. Consequently, this “macro perspective” on an entire administration level could be enriched by further

research from a “meso perspective”, taking into account, for example, specific ministries, and comparing the mechanisms which make a ministry use performance information for a certain purpose, as opposed to other ministries.

Finally yet importantly, this research shows how individual actions and organisational outputs are closely interlinked. It would thus be intriguing to conduct further research focusing on individual behaviour in a group setting. An interesting route to follow could be, for example, ethnographic research, by closely following the “people under study” and taking part in meetings qualified as “learning forums” (or in those that are not, to observe the different mechanisms at play). Obviously, the appropriate language skills are important for such an endeavour.

7.3. Practical implications

The inevitable question researchers of all fields, and in social science in particular, have to deal with is the “so what?” question. As stated in the beginning, the idea underlying this dissertation was a *phronetic* approach to social science, which puts the “so what” question centre stage. Consequently, this section is constructed according to the four value-rational questions raised in the beginning:

- 1) Where are we going with performance management systems?
- 2) Who gains and who loses, by which mechanisms of power?
- 3) Is this development desirable?
- 4) What should we do, if anything?

Question 1 Where are we going with performance management systems?

The title of this thesis considers performance management systems to be “at a crossroads”. When we examine both the literature reviewed and the empirical findings, this seems to be an adequate description. On the one hand, there is abundant evidence available for the negative or unintended effects of using performance information for accountability purposes. Consequently, several authors suggest a change of direction towards using performance information for learning purposes, and holding organisations to account for their learning efforts (Courty & Marschke, 2007; Greiling & Halachmi, 2013; Halachmi, 2002; Kettl, 2001; Mayne, 2007; Molleman & Timmerman, 2003; Olejniczak & Newcomer, 2014; Schillemans et al., 2013; Van Dooren & Willems, forthcoming; Yang & Holzer, 2006). On the other hand, empirical evidence for such a change of direction is scarce. The present study wanted to contribute to analysing performance management systems empirically, and specifically the way performance information is used.

The resulting picture is far from black and white. In Finland, it seems that they have realised that focusing excessively on accountability will lead to a dead end. However, changing direction requires more than just turning the steering wheel. The findings show that the organisational culture is a decisive part of such a change, if not the most decisive one. This is in line with the literature, where, for example, Schick (2003, p. 83)

claims that it is not enough to have performance information available to transform an organisation. In fact, he asserts the following:

“This optimism is rarely justified, for organisations – both public and private – have enormous capacity to assimilate or deflect data on results without changing their policies or operations. It requires sustained political and managerial will to re-orient an organisation in response to information on what it is doing or hopes to accomplish. In fact, genuine organisational change may be a precondition for effective use of performance information.”

Hence, political will and managerial leadership will play an important role for continuing to move towards learning systems (Moynihan & Landuyt, 2009; Moynihan & Lavertu, 2012). As we have seen in the empirical section, some managers are more open to different or new approaches than others are.

In addition, the Lithuanian system is at a crossroads. Since its introduction, the performance management system has constantly been reformed and extended. However, it also seems that with every reform the system became more rigid, which – as warned by Smith (1995) 20 years ago – can lead to organisational paralysis. This appears to be happening in Lithuania. Rather than moving in one clear direction, Lithuania seems to be spinning in circles. Recent reforms and initiatives attempted to further streamline the system, but this created an overflow of indicators, and the cascading of performance targets has been perceived as taking away individual organisations’ room for manoeuvre. The more targets are introduced, the more reporting is required, and the more accountability elements prevail. This in turn leads to more dissatisfaction with the system and distrust among employees. The lack of a shared vision and a clear purpose of the system aggravate the situation. At the same time, Lithuania receives a significant share of its money from the European Structural Funds. In this context, many of the elements of the performance management system are also “externally” required to provide adequate *ex ante* assessments and other reports to the external account-holders. Nevertheless, in theory it is recognised that a move towards more learning elements and a lighter, less rigid, and more flexible system would reduce much of the administrative burden and headache among strategic planning experts. However, it seems that these progressive efforts run counter to the prevailing organisational culture. It remains to be seen whether there will be a way out of this circle.

Question 2 Who gains and who loses; by which mechanisms of power?

The answer to the question of who gains and who loses depends on our departure point. In the current situation, accountability-based systems more closely resemble a lose-lose situation. Those account-givers feel overwhelmed by their reporting duties and, at the same time, question the purpose of this system if the information is clearly not used. To avoid negative reputation or judgement or even sanctions, they set the targets of their performance indicators at “convenient” levels. Furthermore, if positive performance is not “visibly” appreciated, or even discouraged in cases of overachievement resulting in budget cuts to increase efficiency further, account-giving organisations do not feel any

incentive to truly improve their performance. In turn, this does not provide the account-holders with an accurate picture of the organisation at stake. The whole story about setting performance targets and managing performance becomes a farce.

However, the experiences with learning forums and (horizontal) cooperation also show that a different model is possible. When learning elements prevail in a performance management system, it is hoped that this will lead to a win-win situation. Account-givers would discover how they perform; they would feel encouraged to critically discuss their successes and under-performances without being blamed outright. This would contribute to a more collaborative environment and strengthen trust-relationships among employees. People would experience “ownership” of their tasks and policies, feel part of “the bigger picture”, and as such, have a clearer vision of the purpose and the goal of the organisation and their own contribution to it. For account-holders, this would result in acquiring more reliable and meaningful information and having more committed and motivated employees. Ultimately, this would result in better performance.

A critical remark might be appropriate here. This scenario might imply that organisational learning automatically leads to improvement. To a certain extent, this is the underlying assumption of the organisational learning and the learning organisation literature consulted in the scope of this dissertation. Nevertheless, in reality this might not be the case. However, de Lancer Julnes (2008, p. 73) suggests that this is no reason to question learning endeavours and depart from this road.

“[E]ven if the organisation has learned from performance measurement information, program learning may or may not lead to improvement. Instead it may contribute to a better understanding about the program, which may in turn lead to a more informed dialogue. That, too, is a useful contribution of performance measurement.”
(de Lancer Julnes, 2008, p. 73).

This leads us to the second part of the question; what are the mechanisms of power at play? In the lose-lose scenario, the power mechanisms centre on accountability and blame-avoidance. Led by the motivation to avoid blame and negative judgements, organisations adapt their behaviour. The consequences of this action are outlined extensively and do not need to be repeated here. In the win-win scenario, power rests in dialogue and cooperation among employees. It remains to be seen which “forces” gain the upper hand. The answer to the following question depends largely on these mechanisms.

Question 3 Is this development desirable?

The literature reviewed for this dissertation and the empirical investigations demonstrate that using performance information in a “hard” way and, as such, mostly for accountability, is not a desirable development.

However, also so-called “frontrunners” of performance management such as the UK and the Netherlands have in recent years started to question its rationale and have changed direction. In the UK, the Audit Commission was abolished in 2012 as it had been considered “too heavy, burdensome and bureaucratic” (Timmins & Gash, 2014). The idea

was to “pass power down to people, replace bureaucratic accountability with democratic accountability” (Department for Communities & Local Government, 2010). Nevertheless, it is recognised that the abolishment of the Audit Commission has rather been a change in discourse (and a matter of particular political interests) – practice shows that the functions of the Audit Commission will likely be carried out by other institutions, such as “unaccountable private audit firms” (Colin Talbot, 2010; Travers, 2010).

In the Netherlands, there has been a move away from performance-based budgeting towards “accountable budgeting”, or literally translated, “responsible budgeting”, introduced in 2013 (de Jong, van Beek & Posthumus, 2013). Interestingly, in the scope of this reform, about 50 % of the performance indicators included in the budget documents earlier ceased to be in use. This was because “performance information in the old budgets had become more aimed at legitimising funding and compliance than in providing useful insights for oversight or to learn and improve.” To emphasise the latter purpose, focus shifted towards lessons from evaluation in the budget documents, as learning is unlikely to happen “in a cyclical, annual way” (de Jong et al., 2013, p. 2).

Thus, if the road towards learning systems is taken, this could be a desirable development. At the same time, this step is still in its infancy and empirical evidence scarce. Nevertheless, positive and encouraging examples do exist on how to implement such a system in practice.

On the other hand, the question about desirability can be brought down to a fundamental level. Returning to the definition of performance in terms of public values, the question whether this development (or another) is desirable could also be addressed from this point of view, i.e. by asking about the values of public administration one system or the other represents. Then we would have to deal with questions, such as, what kind of public administration is desired? What is a “good” public administration? Why is a lose-lose situation bad? Do only results count, no matter the process? Alternatively, are we also interested in how these processes work? These are fundamental ethical questions showing that the quality of an administration cannot only be judged by the quality of its functions, but that it also depends on the values and intentions to which an administration is orientated.

Question 4 What should be done?

This question can be addressed both from a researcher’s and from a practitioner’s perspective. I provide some suggestions for both perspectives in turn.

From a researcher’s perspective, as already alluded to a few times, more empirical research into the mechanisms of the use of performance information, and in particular, in (potential) learning systems, is necessary. This could include organisational ethnography approaches or participatory action research to get to the “core” where learning might take place (see e.g. Ybema, Yanow, Wels & Kamsteeg, 2009). But also “accountability for learning” models deserve and need more attention and empirical evidence. Furthermore, the empirical findings have shown that organisational culture, leadership, political interests, and trust play an important role. It could thus be

interesting to examine “neighbouring” research disciplines and to enrich the public administration literature with insights from areas such as evaluation (Chelimsky, 2006; Rogers & Williams, 2006), accounting (Brignall & Modell, 2000; Modell, 2009), or human resource management (and public service motivation) (Boselie, Dietz & Boon, 2005; Vermeeren, Kuipers & Steijn, 2014).

From a practitioner’s perspective, several roads are viable. As the present study shows, investment in an organisational culture supportive to cooperation, trust, and trial and error feedback cycles is not a luxury. On the contrary, in a trusting environment, where people share a vision about the goals and purpose of their organisation, where they are not afraid to experiment with new ideas and approaches, and where they are appreciated for such a behaviour, performance management could be elevated to a new and unseen level.

Regarding accountability and supervision, the organisation theorist Günther Ortmann also calls for a change in the approach towards supervision. Focus should shift to supporting individual actors in their self-reflexion and interaction among each other, but also with the cooperative actor organisation – with its structures, functional requirements, and system imperatives (Ortmann, 2011). This implies a revision of the “actor versus system” attitude that manifests itself primarily in blame- and responsibility-shifting. Instead, the mutual dependence of the two must be recognised and consequently nourished. Thus, organisations must engage in codifying values internally and among their members.

This brings us back to the conceptual model underlying the learning discourse, that structure and culture shape actions and behaviour of individuals (and consequently, of the organisation), and at the same time, actors shape structure and culture. This reciprocity must be taken into account when attempting to create understanding and possible improvement or change.

This thesis does not provide a ready-to-use recipe to move towards learning systems. Rather, it provides information about how such a system could look, what the main ingredients could be, and what its advantages would be, and it tested these ideas empirically. In line with the *phronetic* research approach, the contribution of this thesis and the practicalities of the suggestions have to be found in practice through practice. If this thesis and its findings can help change ways towards the road (still) less travelled by, all the better.

8. BIBLIOGRAPHY

- Al-Ani, A. (2015, October 12). Das Prinzip organisierte Unverantwortlichkeit [The principle of organised irresponsibility]. *Die Zeit*. Hamburg.
- Alcock, P. (2004). Targets, indicators and milestones. *Public Management Review*, 6(2), 211–227. <http://doi.org/10.1080/1471903042000189100>
- Ammons, D. N. (2002). Performance measurement and managerial thinking. *Public Performance & Management Review*, 25(4), 344–347. <http://doi.org/10.2307/3381126>
- Argyris, C. & Schön, D. A. (1978). *Organizational learning*. Reading, Mass: Addison-Wesley Pub. Co.
- Askim, J., Johnsen, Å. & Christophersen, K.-A. (2008). Factors behind organizational learning from benchmarking: Experiences from Norwegian municipal benchmarking networks. *Journal of Public Administration Research and Theory*, 18(2), 297–320. <http://doi.org/10.1093/jopart/mum012>
- Aucoin, P. & Heintzman, R. (2000). The dialectics of accountability for performance in public management reform. *International Review of Administrative Sciences*, 66(1), 45–55. <http://doi.org/10.1177/0020852300661005>
- Behn, R. D. (2001). *Rethinking democratic accountability*. Washington, D.C: Brookings Institution Press.
- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586–606. <http://doi.org/10.1111/1540-6210.00322>
- Behn, R. D. & Kant, P. A. (1999). Strategies for avoiding the pitfalls of performance contracting. *Public Productivity & Management Review*, 22(4), 470–489. <http://doi.org/10.2307/3380931>
- Berman, E. (2002). How useful is performance measurement? *Public Performance & Management Review*, 25(4), 348–351. <http://doi.org/10.2307/3381127>
- Berman, E. & Wang, X. (2000). Performance measurement in U.S. counties: Capacity for reform. *Public Administration Review*, 60(5), 409–420.
- Bevan, G. & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84(3), 517–538. <http://doi.org/10.1111/j.1467-9299.2006.00600.x>
- Bijlsma-Frankema, K. (2005). Understanding the trust-control nexus. *International Sociology*, 20(3), 259–282. <http://doi.org/10.1177/0268580905055477>

- Blinder, A. (2015, April 1). Atlanta educators convicted in school cheating scandal. *The New York Times*. New York. Retrieved from http://www.nytimes.com/2015/04/02/us/verdict-reached-in-atlanta-school-testing-trial.html?_r=2
- Bohte, J. & Meier, K. J. (2000). Goal displacement: Assessing the motivation for organizational cheating. *Public Administration Review*, *60*(2), 173–182. <http://doi.org/10.1111/0033-3352.00075>
- Boselie, P., Dietz, G. & Boon, C. (2005). Commonalities and contradictions in HRM and performance research. *Human Resource Management Journal*, *15*(3), 67–94. <http://doi.org/10.1111/j.1748-8583.2005.tb00154.x>
- Bouckaert, G. (2012). Trust and public administration. *Administration*, *60*(1), 91–115.
- Bouckaert, G. & Balk, W. (1991). Public productivity measurement: diseases and cures. *Public Productivity & Management Review*, *15*(2), 229–235. <http://doi.org/10.2307/3380763>
- Bouckaert, G. & Halligan, J. (2008). *Managing performance: International comparisons*. London ; New York: Routledge.
- Bouckaert, G. & Peters, B. G. (2002). Performance measurement and management: The Achilles' heel in administrative modernization. *Public Performance & Management Review*, *25*(4), 359–362. <http://doi.org/10.2307/3381129>
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, *13*(4), 447–468. <http://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Bovens, M., Schillemans, T. & 't Hart, P. (2008). Does public accountability work? An assessment tool. *Public Administration*, *86*(1), 225–242. <http://doi.org/10.1111/j.1467-9299.2008.00716.x>
- Brignall, S. & Modell, S. (2000). An institutional perspective on performance measurement and management in the 'new public sector'. *Management Accounting Research*, *11*(3), 281–306. <http://doi.org/10.1006/mare.2000.0136>
- Burgoyne, J. & Jackson, B. (1997). The arena thesis: Management development as a pluralistic meeting point. In J. Burgoyne & M. Reynolds (Eds.), *Management learning: integrating perspectives in theory and practice* (pp. 54–70). London: SAGE.
- Caldwell, R. (2012). Systems thinking, organizational change and agency: A practice theory critique of Senge's learning organization. *Journal of Change Management*, *12*(2), 145–164. <http://doi.org/10.1080/14697017.2011.647923>
- Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *Handbook of evaluation: Policies, programs and practices* (pp. 33–55). London ; Thousand Oaks, Calif: SAGE.

- Common, R. (2004). Organisational learning in a political environment. *Policy Studies*, 25(1), 35–49.
- Coopey, J. (1995). The learning organization. Power, politics and ideology introduction. *Management Learning*, 26(2), 193–213. <http://doi.org/10.1177/135050769502600204>
- Coopey, J. & Burgoyne, J. (2000). Politics and organizational learning. *Journal of Management Studies*, 37(6), 869–886. <http://doi.org/10.1111/1467-6486.00208>
- Costa, A. C. & Bijlsma-Frankema, K. (2007). Trust and control interrelations: New perspectives on the trust control nexus. *Group & Organization Management*, 32(4), 392–406. <http://doi.org/10.1177/1059601106293871>
- Courty, P. & Marschke, G. (2007). Making government accountable: Lessons from a federal job training program. *Public Administration Review*, 67(5), 904–916. <http://doi.org/10.1111/j.1540-6210.2007.00777.x>
- Curristine, T. (2006). Performance information in the budget process: Results of the OECD 2005 questionnaire. *OECD Journal on Budgeting*, 5(2), 87–131. <http://doi.org/10.1787/budget-v5-art13-en>
- Dahler-Larsen, P. (2014). Constitutive effects of performance indicators: Getting beyond unintended consequences. *Public Management Review*, 16(7), 969–986. <http://doi.org/10.1080/14719037.2013.770058>
- de Bruijn, H. (2002). Performance measurement in the public sector: Strategies to cope with the risks of performance measurement. *International Journal of Public Sector Management*, 15(7), 578–594. <http://doi.org/10.1108/09513550210448607>
- de Jong, M., van Beek, I. & Posthumus, R. (2013). Introducing accountable budgeting: Lessons from a decade of performance-based budgeting in the Netherlands. *OECD Journal on Budgeting*, 12(3), 1–34. <http://doi.org/10.1787/budget-12-5k455r12vs37>
- De Koning, B. (2014, November 11). Opheldering verzocht: acht op de tien misdrijven in Nederland wordt nooit opgelost. *De Correspondent*. Retrieved from <https://decorrespondent.nl/2038/Opheldering-verzocht-acht-op-de-tien-misdrijven-in-Nederland-wordt-nooit-opgelost/230576281342-798134c4>
- De Koning, B. (2015, April 17). De politie is er maar voor de vorm (en dat is maar goed ook). *De Correspondent*. Retrieved from <https://decorrespondent.nl/2725/De-politie-is-er-vooral-voor-de-vorm-en-dat-is-maar-goed-ook/308302437025-bb091a23>
- de Lancer Julnes, P. (2008). Performance measurement beyond instrumental use. In W. Van Dooren & S. Van de Walle (Eds.), *Performance information in the public sector. How it is used*. (pp. 62–75). Basingstoke: Palgrave Macmillan.
- Department for, Communities and, & Local Government. (2010, August 13). Eric Pickles to disband Audit Commission in new era of town hall transparency. Retrieved

from <https://www.gov.uk/government/news/eric-pickles-to-disband-audit-commission-in-new-era-of-town-hall-transparency>

- Die Zeit. (2015, November 4). Grüne sehen Verkehrsminister Dobrindt in der Pflicht [Greens call for Dobrindt, Minister of Transport, to fulfill his duty]. *Die Zeit*.
- Dodge, J., Ospina, S. M. & Foldy, E. G. (2005). Integrating rigor and relevance in public administration scholarship: The contribution of narrative inquiry. *Public Administration Review*, 65(3), 286–300.
- Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, 28(3), 376–417. <http://doi.org/10.2307/3381159>
- Easterby-Smith, M., Crossan, M. & Nicolini, D. (2000). Organizational learning: Debates past, present and future. *Journal of Management Studies*, 37(6), 783–796. <http://doi.org/10.1111/1467-6486.00203>
- Easterby-Smith, M. & Lyles, M. A. (Eds.). (2011a). *Handbook of organizational learning and knowledge management* (2nd ed). Chichester, West Sussex: Wiley.
- Easterby-Smith, M. & Lyles, M. A. (2011b). The evolving field of organizational learning and knowledge management. In M. Easterby-Smith & M. A. Lyles (Eds.), *Handbook of organizational learning and knowledge management* (2nd ed). Chichester, West Sussex: Wiley.
- Ebrahim, A. (2005). Accountability myopia: Losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly*, 34(1), 56–87. <http://doi.org/10.1177/0899764004269430>
- European Commission. (2013a). *Standard Eurobarometer 79 / Spring 2013. Factsheet Finland*. Brussels.
- European Commission. (2013b). *Standard Eurobarometer 79 / Spring 2013. Factsheet Lithuania*. Brussels.
- European Commission. (2015a). *Standard Eurobarometer 83 / Spring 2015. Factsheet Finland*. Brussels.
- European Commission. (2015b). *Standard Eurobarometer 83 / Spring 2015. Factsheet Lithuania*. Brussels.
- Flinders, M. (2011). Daring to be a Daniel. The pathology of politicized accountability in a monitory democracy. *Administration & Society*, 43(5), 595–619. <http://doi.org/10.1177/0095399711403899>
- Flyvbjerg, B. (2001). *Making social science matter. Why social inquiry fails and how it can succeed again*. Oxford, UK; New York: Cambridge University Press.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. <http://doi.org/10.1177/1077800405284363>

- Flyvbjerg, B., Landman, T. & Schram, S. (2012a). Important next steps in phronetic social science. In B. Flyvbjerg, T. Landman, & S. Schram (Eds.), *Real Social Science* (pp. 285–297). Cambridge: Cambridge University Press.
- Flyvbjerg, B., Landman, T. & Schram, S. (Eds.). (2012b). *Real social science. Applied phronesis*. Cambridge: Cambridge University Press. Retrieved from <http://ebooks.cambridge.org/ref/id/CBO9780511719912A011>
- Fryer, K., Antony, J. & Ogden, S. (2009). Performance management in the public sector. *International Journal of Public Sector Management*, 22(6), 478–498. <http://doi.org/10.1108/09513550910982850>
- Furnham, A. (2004). *Management and myths: Challenging business fads, fallacies and fashions*. Basingstoke, Hampshire [UK]; New York: Palgrave Macmillan.
- Garvin, D. A. (1993). Building a learning organisation. *Harvard Business Review*, 71(4), 78–91.
- George, A. & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, Mass: MIT Press.
- Gerring, J. (2007). Is there a (viable) crucial-case method? *Comparative Political Studies*, 40(3), 231–253. <http://doi.org/10.1177/0010414006290784>
- Gilson, C., Dunleavy, P. & Tinkler, J. (2009). *Organizational learning in government sector organizations: Literature review*. London: LSE Public Policy Group (PPG).
- Glaser, S. (1997). Management duckspeak. *Management Decision*, 35(9), 653–655. <http://doi.org/10.1108/00251749710186478>
- Goddard, M. & Mannion, R. (2004). The role of horizontal and vertical approaches to performance measurement and improvement in the UK public sector. *Public Performance & Management Review*, 28(1), 75–95. <http://doi.org/10.1080/15309576.2004.11051822>
- Government of the Republic of Lithuania. (2002, 2010). Strategic Planning Methodology. Resolution No 827, 6 June 2002, amended 25 August 2002, No 1220, Valstybės Žinios (Official Gazette), 2010, No 102-5279 (28-08-2010).
- Government of the Republic of Lithuania. (2012). Public Governance Improvement Programme for 2012 – 2020.
- Gregory, R. & Lonti, Z. (2008). Chasing shadows? Performance measurement of policy advice in New Zealand government departments. *Public Administration*, 86(3), 837–856. <http://doi.org/10.1111/j.1467-9299.2008.00737.x>
- Greiling, D. & Halachmi, A. (2013). Accountability and organizational learning in the public sector. *Public Performance & Management Review*, 36(3), 380–406. <http://doi.org/10.2753/PMR1530-9576360301>

- Grievies, J. (2008). Why we should abandon the idea of the learning organization. *The Learning Organization*, 15(6), 463–473.
<http://doi.org/10.1108/09696470810907374>
- Grizzle, G. A. (2002). Performance measurement and dysfunction: The dark side of quantifying work. *Public Performance & Management Review*, 25(4), 363–369.
<http://doi.org/10.2307/3381130>
- Guijt, I. (2010). Exploding the myth of incompatibility between accountability and learning. In J. Ubels, N.-A. Acquaye-Baddoo, & A. Fowler (Eds.), *Capacity development in practice* (pp. 277–291). London ; Washington, DC: Earthscan.
- Halachmi, A. (2002). Performance measurement, accountability, and improved performance. *Public Performance & Management Review*, 25(4), 370–374.
<http://doi.org/10.2307/3381131>
- Hatry, H. (2011). The many faces of use. In W. Van Dooren & S. Van de Walle (Eds.), *Performance information in the public sector: How it is used*. Basingstoke: Palgrave Macmillan.
- Hedberg, B. (1981). How organizations learn and unlearn. In P. C. Nystrom & W. H. Starbuck (Eds.), *Handbook of Organizational Design. Vol. 1* (pp. 3–27). New York: Oxford 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), 363–394.
- Hirschman, A. O. (1991). *The rhetoric of reaction: perversity, futility, jeopardy*. Cambridge, Mass: Harvard University Press.
- Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed). Thousand Oaks, CA: Sage Publications.
- Hood, C. (1974). Administrative diseases: Some types of dysfunctionality in administration. *Public Administration*, 52(4), 439–454.
<http://doi.org/10.1111/j.1467-9299.1974.tb00196.x>
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19. <http://doi.org/10.1111/j.1467-9299.1991.tb00779.x>
- Hood, C. (2006). Gaming in targetworld: The targets approach to managing British public services. *Public Administration Review*, 66(4), 515–521.
<http://doi.org/10.1111/j.1540-6210.2006.00612.x>
- Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review*, 9(2), 191–210. <http://doi.org/10.1080/14719030701340275>
- Hood, C. (2012). Public management by numbers as a performance-enhancing drug: Two hypotheses. *Public Administration Review*, 72(s1), S85–S92.
<http://doi.org/10.1111/j.1540-6210.2012.02634.x>

- Johansson, V. (2015). When will we ever learn? Presented at the Trans-European Dialogue (TED) 8, Milan, Italy.
- Johansson, V. & Montin, S. (2014). What if performance accountability mechanisms engender distrust? *Urban Research & Practice*, 7(2), 213–227. <http://doi.org/10.1080/17535069.2014.910925>
- Johnsen, A. (1999). Implementation mode and local government performance measurement: a Norwegian experience. *Financial Accountability & Management*, 15(1), 41–66. <http://doi.org/10.1111/1468-0408.t01-1-00073>
- Jung, C. S. & Lee, G. (2013). Goals, strategic planning, and performance in government agencies. *Public Management Review*, 15(6), 787–815. <http://doi.org/10.1080/14719037.2012.677212>
- Kelman, S. & Friedman, J. N. (2009). Performance improvement and performance dysfunction: An empirical examination of distortionary impacts of the emergency room wait-time target in the English national health service. *Journal of Public Administration Research and Theory*, 19(4), 917–946. <http://doi.org/10.1093/jopart/mun028>
- Kerpershoek, E., Groenleer, M. & de Bruijn, H. (2014). Unintended responses to performance management in Dutch hospital care: Bringing together the managerial and professional perspectives. *Public Management Review*, 1–20. <http://doi.org/10.1080/14719037.2014.985248>
- Kettl, D. (2001). Putting performance management to work in the federal government. *La Follette Policy Report*, 12(1), 8–10.
- Kettl, D. F. (2002). *The Transformation of Governance*. Baltimore: John Hopkins University Press.
- Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage Publications.
- Laegreid, P., Roness, P. G. & Rubecksen, K. (2008). Performance information and performance steering: Integrated system or loose coupling? In W. Van Dooren & S. Van de Walle (Eds.), *Performance information in the public sector. How it is used*. (pp. 46–61). Basingstoke: Palgrave Macmillan.
- Levy, J. (2008). Case studies: Types, designs, and logics of inference. *Conflict Management and Peace Science*, 25(1), 1–18. <http://doi.org/10.1080/07388940701860318>
- Lewis, J. M. & Triantafillou, P. (2012). From performance measurement to learning: A new source of government overload? *International Review of Administrative Sciences*, 78(4), 597–614. <http://doi.org/10.1177/0020852312455993>
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications.

- Lipshitz, R., Popper, M. & Oz, S. (1996). Building learning organizations: The design and implementation of organizational learning mechanisms. *The Journal of Applied Behavioral Science*, 32(3), 292–305. <http://doi.org/10.1177/0021886396323004>
- Luton, L. S. (2010). *Qualitative research approaches for public administration*. Armonk, N.Y: M.E. Sharpe, Inc.
- Mahler, J. (1997). Influences of organizational culture on learning in public agencies. *Journal of Public Administration Research and Theory*, 7(4), 519–540.
- Marr, B. (2009). *Managing and delivering performance: How government, public sector and not-for-profit organizations can measure and manage what really matters*. Amsterdam; Boston: Butterworth-Heinemann/Elsevier.
- Mayne, J. (2007). Evaluation for accountability: Myth or reality? In M.-L. Bemelmans-Videc, J. Lonsdale, & B. Perrin (Eds.), *Making accountability work: dilemmas for evaluation and for audit* (pp. 63–84). New Brunswick: Transaction Publishers.
- Ministry of Finance. (2006). *Handbook on performance management*. Helsinki: Ministry of Finance.
- Ministry of Finance. (2010). *Finnish public governance - A background report* (No. 18/2010). Helsinki: Ministry of Finance.
- Ministry of Finance. (2012). *Towards more strategic, lighter, more cross-administrative and more integrated performance management [Kohti strategisempää, kevyempää, poikkihallinnollisempää ja yhtenäisempää tulosohtausta] Final report of the Performance Management Development Project* (No. 21/2012). Helsinki.
- Ministry of Finance. (2013a). *Governments of the future. Main report November 2013*. Helsinki, Finland.
- Ministry of Finance. (2013b). *Short description of performance management/results management in Finland*. Helsinki, Finland: Ministry of Finance, Finland.
- Modell, S. (2009). Institutional research on performance measurement and management in the public sector accounting literature: A review and assessment. *Financial Accountability & Management*, 25(3), 277–303. <http://doi.org/10.1111/j.1468-0408.2009.00477.x>
- Molleman, E. & Timmerman, H. (2003). Performance management when innovation and learning become critical performance indicators. *Personnel Review*, 32(1), 93–113. <http://doi.org/10.1108/00483480310454745>
- Möllering, G. (2005). The trust/control duality: An integrative perspective on positive expectations of others. *International Sociology*, 20(3), 283–305. <http://doi.org/10.1177/0268580905055478>
- Morieux. (2011). Smart rules: Six ways to get people to solve problems without you. *Harvard Business Review*, 89(9), 78–84.

- Morieux, Y. (2015). *How too many rules at work keep you from getting things done [Video file]*. BCG London. Retrieved from https://www.ted.com/talks/yves_morieux_how_too_many_rules_at_work_keep_you_from_getting_things_done
- Moynihan, D. (2005). Goal-based learning and the future of performance management. *Public Administration Review*, 65(2), 203–216.
- Moynihan, D. (2011). Advocacy and learning: An interactive-dialogue approach to performance information use. In W. Van Dooren & S. Van de Walle (Eds.), *Performance information in the public sector: How it is used* (pp. 28–45). Basingstoke: Palgrave Macmillan.
- Moynihan, D. (2014, June 1). The problem at the VA: ‘Performance perversity’. *Los Angeles Times*. Retrieved from <http://www.latimes.com/opinion/op-ed/la-oe-moynihan-va-scandal-performance-perversity-20140602-story.html>
- Moynihan, D. & Landuyt, N. (2009). How do public organizations learn? Bridging cultural and structural perspectives. *Public Administration Review*, 69(6), 1097–1105. <http://doi.org/10.1111/j.1540-6210.2009.02067.x>
- Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Washington, D.C: Georgetown University Press.
- Moynihan, D. P. (2009). Through a glass, darkly: Understanding the effects of performance regimes. *Public Performance & Management Review*, 32(4), 592–603. <http://doi.org/10.2753/PMR1530-9576320409>
- Moynihan, D. P., Fernandez, S., Kim, S., LeRoux, K. M., Piotrowski, S. J., Wright, B. E. & Yang, K. (2011). Performance Regimes Amidst Governance Complexity. *Journal of Public Administration Research and Theory*, 21(suppl 1), i141–i155. <http://doi.org/10.1093/jopart/muq059>
- Moynihan, D. P. & Hawes, D. P. (2012). Responsiveness to reform values: The influence of the environment on performance information use. *Public Administration Review*, 72(s1), S95–S105. <http://doi.org/10.1111/j.1540-6210.2012.02653.x>
- Moynihan, D. P. & Lavertu, S. (2012). Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART. *Public Administration Review*, 72(4), 592–602. <http://doi.org/10.1111/j.1540-6210.2011.02539.x>
- Mulgan, R. G. (2003). *Holding power to account: Accountability in modern democracies*. Houndmills, Basingstoke, Hampshire ; New York: Palgrave Macmillan.
- Nakrošis, V. (2001). Lithuanian public administration: A usable state bureaucracy? *Journal of Baltic Studies*, 32(2), 170–181. <http://doi.org/10.1080/01629770100000041>
- Nakrošis, V. (2008). Reforming performance management in Lithuania: Towards results-based government. In B. G. Peters (Ed.), *Mixes, matches, and mistakes: new*

- public management in Russia and the former Soviet republics* (pp. 53–115). Budapest: Local Government and Public Service Reform Initiative, Open Society Institute.
- Nielsen, P. A. (2014). Performance management, managerial authority, and public service performance. *Journal of Public Administration Research and Theory*, 24(2), 431–458. <http://doi.org/10.1093/jopart/mut025>
- Nielsen, P. A. & Baekgaard, M. (2015). Performance information, blame avoidance, and politicians' attitudes to spending and reform: Evidence from an experiment. *Journal of Public Administration Research and Theory*, 25(2), 545–569. <http://doi.org/10.1093/jopart/mut051>
- Nõmm, K. & Randma-Liiv, T. (2012). Performance measurement and performance information in new democracies. *Public Management Review*, 14(7), 859–879. <http://doi.org/10.1080/14719037.2012.657835>
- Noordegraaf, M. (2008). Meanings of measurement: The real story behind the Rotterdam Safety Index. *Public Management Review*, 10(2), 221–239. <http://doi.org/10.1080/14719030801928672>
- Noordegraaf, M. & Abma, T. (2003). Management by measurement? Public management practices amidst ambiguity. *Public Administration*, 81(4), 853–871. <http://doi.org/10.1111/j.0033-3298.2003.00374.x>
- Norman, R. (2001). Letting and making managers manage: the effect of control systems on management action in New Zealand's central government. *International Public Management Journal*, 4(1), 65–89. [http://doi.org/10.1016/S1096-7494\(01\)00043-5](http://doi.org/10.1016/S1096-7494(01)00043-5)
- OECD. (2009). *Measuring government activity*. Paris: Organisation for Economic Co-operation and Development.
- OECD. (2010). *Finland: Working together to sustain success*. Paris: OECD.
- OECD. (2015a). *Lithuania: Fostering open and inclusive policy making*. OECD Publishing.
- OECD. (2015b). Open Government. Retrieved 26 October 2015, from <http://www.oecd.org/gov/open-government.htm>
- Office of the Prime Minister of The Republic of Lithuania. (2011). Methodology for formulating and applying performance measures used in strategic planning documents.
- Olejniczak, K. (2014). Introduction. In K. Olejniczak & S. Mazur (Eds.), *Organizational Learning. A Framework for Public Administration* (pp. 11–19). Warsaw: Scholar Publishing House.
- Olejniczak, K. & Mazur, S. (Eds.). (2014). *Organizational learning. A framework for public administration*. Warsaw: Scholar Publishing House.

- Olejniczak, K. & Newcomer, K. (2014). Moving towards accountability for learning. In K. Olejniczak & S. Mazur (Eds.), *Organizational Learning. A Framework for Public Administration* (pp. 81–98). Warsaw: Scholar Publishing House.
- Olsen, J. & Peters, B. G. (1996). Learning from experience? In *Lessons from experience: Experimental learning in administrative reforms in eight democracies* (pp. 1–35). Oslo: Scandinavian Univ. Press.
- O'Neill, O. (2002). *A question of trust*. Cambridge; New York: Cambridge University Press.
- Oomsels, P. & Bouckaert, G. (2014). Studying interorganizational trust in public administration. *Public Performance & Management Review*, 37(4), 577–604. <http://doi.org/10.2753/PMR1530-9576370403>
- Oomsels, P., Porumbescu, G., Bouckaert, G. & Im, T. (2013). Less or more control to build trust? Exploring the relationship between institutionalised control and administrative trust in a cultural framework. Presented at the ECPR General conference, 4-7 September, Bordeaux.
- Örtenblad, A. (Ed.). (2013a). *Handbook of research on the learning organization*. Cheltenham, UK: Edward Elgar Publishing.
- Örtenblad, A. (2013b). What do we mean by 'learning organisation'? In *Handbook of Research on the Learning Organization* (pp. 22–34). Cheltenham, UK: Edward Elgar Publishing.
- Ortmann, G. (2011). *Moralverdrängung in und durch Organisationen [Moral displacement in and by organisations]*. Kassel: Kassel University Press. Retrieved from <http://d-nb.info/1018144242/34>
- Osborne, D. E. & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. Reading, Ma: Addison-Wesley.
- Oswick, C., Anthony, P., Keenoy, T., Mangham, I. L. & Grant, D. (2000). A dialogic analysis of organizational learning. *Journal of Management Studies*, 37(6), 887–902. <http://doi.org/10.1111/1467-6486.00209>
- Pidd, M. (2005). Perversity in public service performance measurement. *International Journal of Productivity and Performance Management*, 54(5/6), 482–493. <http://doi.org/10.1108/17410400510604601>
- Pivoras, S. (2008). Post-Soviet transformation of bureaucracy in Lithuania: Main features and trends. *Baltic Journal of Law & Politics*, 1(1), 111–131. <http://doi.org/10.2478/v10076-008-0007-z>
- Pivoras, S. (2013). Post-communist public administration in Lithuania. In S. Liebert, S. E. Condrey, & D. V. Goncharov (Eds.), *Public administration in post-communist countries: Former Soviet Union, Central and Eastern Europe, and Mongolia* (pp. 135–160). Boca Raton: CRC Press.

- Pollitt, C. (1995). Improvement strategies. In C. Pollitt & G. Bouckaert (Eds.), *Quality improvement in European public services: Concepts, cases and commentary* (pp. 131–161). London ; Thousand Oaks, Calif: Sage.
- Pollitt, C. & Bouckaert, G. (2011a). *Public Management Reform: A comparative analysis - new public management, governance, and the Neo-Weberian state*. Oxford University Press.
- Pollitt, C. & Bouckaert, G. (2011b). *Public management reform: A comparative analysis - New Public Management, Governance, and the Neo-Weberian State*. Oxford: Oxford University Press.
- Pollitt, C. & Hupe, P. (2011). Talking about government: The role of magic concepts. *Public Management Review*, 13(5), 641–658.
<http://doi.org/10.1080/14719037.2010.532963>
- Prime Minister's Office, & Ministry of Finance. (2013). *Decisions or changes: Finland in the mirror. Governments for the Future*. Helsinki, Finland.
- Radin, B. (2000). The Government Performance and Results Act and the tradition of federal management reform: Square pegs in round holes? *Journal of Public Administration Research and Theory*, 10(1), 111–135.
- Radin, B. (2006). *Challenging the performance movement: Accountability, complexity, and democratic values*. Washington, D.C: Georgetown University Press.
- Rainey, H. G. (2009). *Understanding and Managing Public Organizations*. John Wiley & Sons.
- Rashman, L., Withers, E. & Hartley, J. (2009). Organizational learning and knowledge in public service organizations: A systematic review of the literature. *International Journal of Management Reviews*, 11(4), 463–494.
<http://doi.org/10.1111/j.1468-2370.2009.00257.x>
- Republic of Lithuania. Law on Public Administration, Pub. L. No. VIII-1234, 17 June 1999, as last amended on 2 July 2013, No. XII-511 (1999).
- Richter, I. (1998). Individual and organizational learning at the executive level: Towards a research agenda. *Management Learning*, 29(3), 299–316.
<http://doi.org/10.1177/1350507698293003>
- Rimkute, E. (2014). *Managing for results in Lithuania*. Presented at the Studievoormiddag 'Naar een meer resultaatgerichte Vlaamse overheid'. 10 January 2014, Brussels.
- Rogers, P. J. & Williams, B. (2006). Evaluation for practice improvement and organisational learning. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *Handbook of evaluation: Policies, programs and practices* (pp. 76–97). London ; Thousand Oaks, Calif: SAGE.

- Romzek, B. S. (2000). Dynamics of public sector accountability in an era of reform. *International Review of Administrative Sciences*, 66(1), 21–44.
<http://doi.org/10.1177/0020852300661004>
- Saetren, H. (2012, September). *The critical/crucial case method: what, why, how?* Presented at the EGPA PhD Symposium, Bergen, Norway.
- Salminen, M. & Viitala, M.-L. (2006). *Handbook on performance management*. [Helsinki]: Ministry of Finance, Public Management Department : distributed by Edita Publishing.
- Schick, A. (2003). The performing state: Reflection on an idea whose time has come but whose implementation has not. *OECD Journal on Budgeting*, 3(2), 71–103.
<http://doi.org/10.1787/budget-v3-art10-en>
- Schillemans, T. (2008). Accountability in the shadow of hierarchy: The horizontal accountability of agencies. *Public Organization Review*, 8(2), 175–194.
<http://doi.org/10.1007/s11115-008-0053-8>
- Schillemans, T. & Smulders, R. (2016). Learning from accountability?! Whether, what, and when. *Public Performance & Management Review*, 39(1), 248–271.
<http://doi.org/10.1080/15309576.2016.1071175>
- Schillemans, T., Twist, M. V. & Vanhomerig, I. (2013). Innovations in accountability. *Public Performance & Management Review*, 36(3), 407–435.
<http://doi.org/10.2753/PMR1530-9576360302>
- Schulz, B. (2015, November 5). Faked in Germany. *Die Zeit*.
- Senge, P. M. (2006). *The fifth discipline: The art and practice of the learning organization* (2nd ed). New York, N.Y.; London: Currency Doubleday.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18(2-3), 277–310.
<http://doi.org/10.1080/01900699508525011>
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks: Sage Publications.
- Sterck, M. (2007). The impact of performance budgeting on the role of the legislature: A four-country study. *International Review of Administrative Sciences*, 73(2), 189–203. <http://doi.org/10.1177/0020852307077960>
- Sustainable Governance Indicators (SGI). (2014). *2014 Lithuania report*. Edited by Vitalis Nakrosis, Ramunas Vilpisauskas, Detlef Jahn (Coordinator). Gütersloh: Bertelsmann Stiftung.
- Talbot, C. (2010, August 13). The Audit Commission - chronicle of a death unforeshadowed. Retrieved from <http://blog.policy.manchester.ac.uk/whitehallwatch/2010/08/from-ms-to-bhs-going-down-market-in-the-efficiency-stakes/>
- Talbot, C. & Talbot, C. (2008). UK parliamentary scrutiny of public service agreements: A challenge too far? In W. Van Dooren & S. Van de Walle (Eds.), *Performance*

- information in the public sector. How it is used.* (pp. 144–160). Basingstoke: Palgrave Macmillan.
- Taylor, J. (2009). Strengthening the link between performance measurement and decision making. *Public Administration*, 87(4), 853–871.
<http://doi.org/10.1111/j.1467-9299.2009.01788.x>
- ter Bogt, H. J. (2004). Politicians in search of performance information? Survey research on Dutch aldermen's use of performance information. *Financial Accountability and Management*, 20(3), 221–252. <http://doi.org/10.1111/j.0267-4424.2004.00387.x>
- Thomas, G. (2010). Doing case study: Abduction not induction, phronesis not theory. *Qualitative Inquiry*, 16(7), 575–582. <http://doi.org/10.1177/1077800410372601>
- Thomas, G. (2011a). *How to do your case study: A guide for students and researchers*. Los Angeles, CA ; London: SAGE.
- Thomas, G. (2011b). The case: Generalisation, theory and phronesis in case study. *Oxford Review of Education*, 37(1), 21–35.
<http://doi.org/10.1080/03054985.2010.521622>
- Timmins, N. & Gash, T. (2014). *Dying to improve. The demise of the Audit Commission and other improvement agencies*. London: Institute for Government. Retrieved from <http://www.instituteforgovernment.org.uk/publications/dying-improve>
- Travers, T. (2010, August 15). Shock and awe as commission axed. Retrieved from <http://www.publicfinance.co.uk/2010/08/shock-and-awe-as-commission-axed-by-tony-travers>
- Vakkuri, J. & Meklin, P. (2006). Ambiguity in Performance Measurement: A Theoretical Approach to Organisational Uses of Performance Measurement. *Financial Accountability & Management*, 22(3), 235–250.
- Van Dooren, W. (2011). Better performance management: Some single- and double-loop strategies. *Public Performance & Management Review*, 34(3), 420–433.
<http://doi.org/10.2753/PMR1530-9576340305>
- Van Dooren, W., Bouckaert, G. & Halligan, J. (2015). *Performance management in the public sector* (2nd ed.). London: Routledge.
- Van Dooren, W., De Caluwe, C. & Lonti, Z. (2012). How to measure public administration performance. *Public Performance & Management Review*, 35(3), 489–508. <http://doi.org/10.2753/PMR1530-9576350306>
- Van Dooren, W. & Willems, T. (forthcoming). Thinking allowed: Reforming accountability to enhance innovation. In P. Triantafillou & J. Torfing (Eds.), *Enhancing innovation by transforming public sector governance*. Cambridge: Cambridge University Press.
- Van Thiel, S. & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance and Management Review*, 25(3), 267–281.

- Vermeeren, B., Kuipers, B. & Steijn, B. (2014). Does leadership style make a difference? Linking HRM, job satisfaction, and organizational performance. *Review of Public Personnel Administration*, 34(2), 174–195.
<http://doi.org/10.1177/0734371X13510853>
- Watkins, K. E. & Golembiewski, R. T. (1995). Rethinking organization development for the learning organization. *International Journal of Organizational Analysis (1993 - 2002)*, 3(1), 86–101.
- Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies*. New York: Free Press.
- Wiggins, A. & Tymms, P. (2002). Dysfunctional effects of league tables: A comparison between English and Scottish primary schools. *Public Money & Management*, 22(1), 43–48. <http://doi.org/10.1111/1467-9302.00295>
- Willems, T. & Van Dooren, W. (2011). Lost in diffusion? How collaborative arrangements lead to an accountability paradox. *International Review of Administrative Sciences*, 77(3), 505–530. <http://doi.org/10.1177/0020852311408648>
- Willems, T. & Van Dooren, W. (2012). Coming to terms with accountability. *Public Management Review*, 14(7), 1011–1036.
<http://doi.org/10.1080/14719037.2012.662446>
- Yang, K. & Holzer, M. (2006). The performance–trust link: Implications for performance measurement. *Public Administration Review*, 66(1), 114–126.
<http://doi.org/10.1111/j.1540-6210.2006.00560.x>
- Ybema, S., Yanow, D., Wels, H. & Kamsteeg, F. H. (Eds.). (2009). *Organizational ethnography: Studying the complexity of everyday life*. Los Angeles ; London: SAGE.
- Zucchini, D., Carcamo, C. & Zarembo, A. (2014, May 18). Growing evidence points to systemic troubles in VA healthcare system. *Los Angeles Times*. Retrieved from <http://www.latimes.com/nation/la-na-va-delays-20140518-story.html>

SAMENVATTING (DUTCH SUMMARY)

Het gebruik van prestatimanagementsystemen bevindt zich momenteel op een kruispunt. De introductie van New Public Management (NPM)-geïnspireerde hervormingen versterkte de administratieve verantwoording. Getuige hiervan is de overvloed aan prestatie-indicatoren die als doel hebben de output of activiteiten van de openbare sector te meten. Dit heeft belangrijke implicaties voor prestatimanagementsystemen. In veel gevallen leidt een te sterke focus op verantwoording tot verschillende negatieve effecten, zoals data- of outputmanipulatie. De literatuur erkent dan ook dat administratieve, indicator-gebaseerde verantwoording haar doel heeft gemist. Tegelijkertijd is verantwoording essentieel voor democratische systemen. Hoe moeten we er dan omgaan?

Wetenschappers schuiven als alternatief de leerdoeleinden van indicatoren naar voor. In plaats van actoren verantwoordelijk te houden voor hun resultaten, moeten ze verantwoording afleggen voor de processen waarbij ze betrokken zijn. De focus ligt op de kwaliteit van handelen, niet zozeer op (mis)lukken hiervan. De hiërarchische, vaak externe verantwoordingsrelaties van actoren worden dan vervangen door betrokkenheid in horizontale, interne leerinspanningen met een gedeelde verantwoordelijkheid voor het handelen. Bewijs dat overheden daadwerkelijk leren van prestatie-informatie is echter schaars.

De vraag rijst of verantwoordings- en leerdoeleinden elkaar uitsluiten of er toch een middenweg is die de twee doeleinden met elkaar verbindt? Empirisch onderzoek hieromtrent is schaars. De *organisational learning* literatuur focuste tot nog toe voornamelijk op de privésector en biedt weinig inzichten in de uitdagingen voor verantwoording in de publieke sector. De literatuur rond verantwoording besteedt dan weer meer aandacht aan prestatimanagementprocedures en minder aan leren. Er bestaat dus weinig kennis over hoe leren en prestatieverantwoording kunnen samengaan.

Dit doctoraatsonderzoek bestudeert hoe overheden meer op leren kunnen inzetten via een case analyse van Finse en Litouwse prestatie-meetsystemen. Ik stel dat prestatiemeting als lerend managementsysteem pas ontwikkeld kan worden wanneer er minder nadruk ligt op verantwoording.

Er zijn verschillende redenen om aan te nemen dat prestatie-informatie die wordt verzameld voor verantwoordingsdoeleinden moeilijk samengaat met leerdoeleinden. De Finse en Litouwse prestatimanagementsystemen bevestigden deze vaststelling. Er is sprake van een overvloed aan indicatoren, die vaak niet gebruikt worden. Dit resulteert tot ontevredenheid bij werknemers die de dataverzameling als een grote last ervaren. Wanneer prestatie-informatie bovendien gekoppeld is aan sancties of beloningen, stijgt

de kans op datamanipulatie. De verkregen data zijn bijgevolg betekenisloos en kunnen niet bijdragen in de discussie.

Als alternatief stel ik een lerend systeem voor waar prestatie-informatie vooral ingezet wordt voor leerdoeleinden. Verantwoording wordt losgekoppeld van prestatie-informatie. Die prestatie-informatie dient voornamelijk als input voor de verbetering van beleidsprocessen en richt de blik op de toekomst. Verantwoording daarentegen kijkt vooral naar het verleden. Een lerend systeem moedigt bovendien interactie tussen alle betrokken stakeholders aan. Dit kan plaatsvinden in zogenaamde leerfora, momenten waarop een dialoog over prestaties kan plaatsvinden. Prestatie-indicatoren kunnen in deze leerfora aan bod komen, aangevuld met informatie uit andere bronnen. De omslag van een verantwoordings- naar een leersysteem vergt zowel structurele als culturele veranderingen.

Mijn onderzoek toont aan dat verantwoordings- en leerdoeleinden moeilijk te combineren zijn. Reden hiervoor is dat ruimte voor interpretatie die verschilt. Een leersysteem gaat uit van dialoog en uitwisseling met interpretatie van de prestatie-informatie. Het maakt gebruik van verschillende informatiebronnen en vult harde feiten aan met kennis gebaseerd op ervaringen. Een verantwoordingsstelsel daarentegen laat geen ruimte voor interpretatie. Zodra informatie extern gepubliceerd wordt, vervalt de mogelijkheid tot interpretatie en discussie in leerfora. De druk stijgt op degene die verantwoording moet afleggen om goede resultaten te laten zien, waardoor het gevaar om de hoek loert dat data 'aangepast' wordt aan de verwachtingen van de toezichhouders. De drempel voor data- of outputmanipulatie ligt redelijk laag, zo blijkt uit de wetenschappelijke literatuur. Informatie die voor verantwoordingsdoeleinden gebruikt wordt, is daarom niet meer betekenisvol voor leerdoeleinden. Sterker nog, mijn onderzoek bevestigt dat informatie die voor verantwoordingsdoeleinden verzameld wordt niet tot betekenisvol prestatie-management leidt. Daarom stel ik dat prestatie-informatie losgekoppeld moet worden van verantwoordingsdoeleinden en beter ingezet kan worden voor leerdoeleinden. Dit kan prestatie-management opnieuw betekenisvol maken.