WHY AGENCIES BUDGET FOR RESULTS

Exploring Institutional Explanations for Performance Budgeting: The Case of Forestry and Air Traffic Control

Maarten de Jong

Colofon

Why Agencies Budget For Results - Exploring Institutional Explanations for Performance Budgeting: The Case of Forestry and Air Traffic Control

PhD thesis Erasmus University Rotterdam

Doctoral Committee:

Promotor(s):
Prof. dr. C.W.A.M. Van Paridon (Erasmus University)

Other members:
Prof. dr. S. Van de Walle (Erasmus University)
Prof dr. S. Van Thiel (Radboud University)
Prof. dr. D.P. Moynihan (University of Wisconsin)

Copromotor(s):
Jhr. dr. F.K.M. Van Nispen tot Pannerden (Erasmus University)

External reader:
Prof. dr. P.L. Posner (George Mason University)

WHY AGENCIES BUDGET FOR RESULTS

Exploring Institutional Explanations for Performance Budgeting:
The Case of Forestry and Air Traffic Control

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Thursday January 14th 2016 at 15.30 hrs
by

Maarten de Jong

SUMMARY

This PhD research examines institutional explanations for the use of performance information (PI) to learn and improve by government agencies. Despite widespread criticism of and disappointment with New Public Management (NPM) reforms, performance budgeting (PB) is credited with contributing to efficiency and effectiveness in several public sector agencies. The fact that an agency successfully uses performance information to realize improvements or efficiency gains may however be the result of a diversity of factors that bear little or no connection to the PB system itself. Corporate culture, public sector motivation, leadership commitment and discretionary power of managers have already been identified as important in this respect. Moreover, the causal relationship between successful PB adoption and the purposeful use of performance information (PI) by an agency may in fact be reversed. The central question posed in this research is therefore: *How do underlying cultural and historical factors explain successful PB in government agencies?*

In an attempt to further explore the 'DNA' of agencies with a reputation of successful performance budgeting, this study turned to alternative explanations provided by neo institutionalism. More specifically, a number of indicators concerning an organization's history and culture were explored as explanations for purposeful use of PI. Among these were: reflective openness, cognitive frames the impact of events and leadership.

The research consisted of four qualitative case studies of four public sector agencies in the Netherlands and the US that share a reputation of purposeful use of performance information, that is, PI use that is consistent with PB reform objectives. Four indicators were used to assess the degree to which a case can indeed be qualified as a PB success. Subsequently specific institutional factors believed to be beneficial for PI use were tested by eight indicators. The primary source of evidence consisted of semi structured interviews with key organization members. Overall, 36 persons were interviewed who were selected to represent management, performance & budget staff and operations of each agency as well as staff tasked with oversight on the part of the agency's principal. By means of triangulation the findings from interviews were compared with questionnaire results and findings from analysis of documents.

The results indicate that in all agencies examined, deeply rooted cultural characteristics largely explain purposeful use of performance information. In two of the cases, events or leaders provide an additional explanation. Although one should be careful with generalization from a limited number of case studies, the results cast doubts on some widely held assumptions among budget reformists. As the cases studied were carefully selected to represent the relatively scarce good practices of PB implementation, the findings challenge the assumption that purposeful use of PI results from PB reforms.

While PB adoption may not provide the main explanation for purposeful performance information use by these agencies, it can be credited with positively impacting effectiveness and efficiency in the areas of goal alignment, capacity planning and stakeholder dialogue. The analysis from the case studies suggests that cultural and historical institutional factors are a necessary conditions for PB to work as envisaged Therefore anecdotal evidence of purposeful PI use by government agencies should not be mistaken for success of a NPM/PB recipe that can be reproduced elsewhere. For future attempts to advance this popular public financial management (PFM) reform, it is suggested that measurement and reporting systems should remain close to organization members' own measures of professionalism and that reforms should simultaneously address aspects of organization culture associated with organizational learning.

SAMENVATTING

Dit promotieonderzoek richt zich op institutionele verklaringen voor het gebruik van prestatie-informatie (PI) door publieke organisaties om te leren en te verbeteren. Er is momenteel veel kritiek op - en teleurstelling over *New Public Management* (NPM) hervormingen. Toch lijkt bij sommige organisaties *performance budgeting* (PB) te hebben bijgedragen aan meer doeltreffendheid en doelmatigheid. Het feit dat een organisatie prestatie-informatie succesvol gebruikt om verbeteringen of doelmatigheidswinst te realiseren kan echter het gevolg zijn van uiteenlopende factoren die geheel los staan van een PB systeem. Factoren als organisatiecultuur, intrinsieke motivatie, toegewijd leiderschap en de mate van autonomie zijn daarbij in eerder onderzoek al als belangrijk geïdentificeerd. Het zou zelfs zo kunnen zijn dat het vaak veronderstelde oorzakelijk verband tussen het succesvol invoeren van PB en het zinvol gebruiken van PI, in feite omgekeerd is. De centrale vraag van dit onderzoek is daarom: *Hoe verklaren onderliggende culturele en historische factoren succesvolle toepassing van PB in publieke organisaties*?

In een poging het 'DNA' te ontrafelen van organisaties met een reputatie van succesvolle PB toepassing, kijkt dit onderzoek naar alternatieve verklaringen gebaseerd op het neo-institutionalisme. Hiertoe wordt een aantal variabelen met betrekking tot de geschiedenis en cultuur van een organisatie getoetst als mogelijke verklaring voor het zinvol gebruik van PI. Deze variabelen zijn onder meer: analytische openheid, gedeelde overtuigingen, de gevolgen van ingrijpende gebeurtenissen en leiderschap.

Het onderzoek bestaat uit vier kwalitatieve case studies van vier publieke organisaties uit Nederland en de VS die allen een reputatie hebben van zinvol gebruik van PI, dat wil zeggen dat zij PI gebruiken op een manier die overeenkomt met de doelstellingen van PB hervormingen. Vier indicatoren meten in hoeverre een case inderdaad kan worden beschouwd als PB succes. Vervolgens wordt met acht indicatoren de aanwezigheid getoetst van specifieke institutionele factoren die als gunstig worden beschouwd voor PI gebruik. De belangrijkste bron voor dit onderzoek zijn semi gestructureerde interviews met relevante organisatieleden. In totaal zijn 36 personen geïnterviewd die een afspiegeling vormen van het management, de uitvoering en de stafdiensten voor prestatie-management & begroting. Ook zijn toezichthouders van de principaal geïnterviewd. Door triangulatie zijn de bevindingen uit de interviews vergeleken met enquêteresultaten en schriftelijke bronnen.

De resultaten wijzen er op dat in al deze organisaties diepgewortelde culturele kenmerken het zinvol gebruik van PI grotendeels verklaren. In twee cases vormen specifieke gebeurtenissen en leiders een plausibele aanvullende verklaring. Hoewel voorzichtigheid is geboden bij generalisatie uit enkele cases, werpen de bevindingen een nieuw licht op een aantal populaire aannames. Aangezien de bestudeerde cases golden al relatief schaarse voorbeelden van PB succes, trekken de resultaten de aanname in twijfel dat zinvol gebruik van PI volgt uit PB hervormingen.

Hoewel PB niet de belangrijkste verklaring vormt van zinvol PI-gebruik door deze organisaties, heeft PB wel positief bijgedragen aan effectiviteit en efficiency door middel van de afstemming van organisatiedoelen, een rationele capaciteitsplanning en de dialoog tussen stakeholders. Uit de analyse van de cases komt naar voren dat culturele en historische institutionele factoren noodzakelijke voorwaarden vormen om PB te laten werken zoals bedoeld. Om die reden is het onverstandig om anekdotisch bewijs van zinvol PI gebruik door overheidsorganisaties aan te zien voor generiek succes van een NPM/PB recept dat elders kan worden gekopieerd. Voor meer succesvolle toepassing van deze populaire begrotingshervorming wordt geadviseerd om meet- en rapportagesystemen te kiezen die aansluiten bij de wijze waarop organisatieleden succes beoordelen volgens hun eigen professionele maatstaven. Ook zouden prestatiegerichte hervormingen zich tegelijkertijd moeten richten op culturele aspecten die van belang zijn voor het lerend vermogen van organisaties.

ACKNOWLEGDEMENTS

common quest for this academic title but I suspect something must have gone wrong in our early youth. On a slightly more serious note I'd like to thank everyone who openly doubted my intellectual capacity. Growing up with a speech impediment, there have been quite a few throughout the years: from my elementary school teachers to my dean in university. Your suggestions regarding my future career or sometimes your conviction of the lack of any, may have inspired me to set more difficult goals for myself at crucial times.

From the early start of this project until the very end, Frans Van Nispen was always there for valuable advice, even in the weekends. Each time, our meetings were longer than anticipated as we spent most of the time discussing fiscal affairs in Europe and The Hague. Paul Posner's support has also been pivotal to this project. Apart from all his knowledge on budgeting, Paul's incredible DC network and knowledge of the US federal government, meant that every door would literally open for me. I recall that while sometimes it took two weeks and long registration forms to get an appointment at some agency building, I was able to enter the White House's Eisenhower Building with only a marginal security check within a few days' notice from Paul. In addition to Frans and Paul I would like to thank Kees van Paridon and Peter van den Berg who, both from their own role, offered great advice to this project. Our meetings were always in a constructive and good humored atmosphere.

I would also like to thank many colleagues at the Ministry of Finance who regularly showed interest by asking about the progress of my research and who also provided me with the much needed flexibility to combine family, research, work and foreign projects at times. This goes in particular my head manager Joost van Hofwegen. At Erasmus University I was formally enrolled as unpaid 'staff' at the faculty of Social Sciences at Erasmus University. As an external PhD candidate I never found the time to fully integrate as a member of the Erasmus community. Nonetheless I enjoyed the positive atmosphere and have always found EUR staff and fellow PhD candidates to be extremely kind and helpful.

Finally, I would like to thank all the other people who made this research possible. Firstly the dedicated public sector professionals in the Netherlands and the US who agreed to be interviewed by me for this research. Also all the people who helped me set up and organize interviews and questionnaires and provided me with other kinds of information. I would like thank everyone who inspired me by discussing this topic or helped me by commenting on my papers or (parts of the) draft manuscripts: Bob Behn, Wouter van Dooren, Matt Dull, Aimee Franklin, Phil Joyce, Henk Klaassen, Julianne Mahler, Don Moynihan, Marc Robinson, Sandra van Thiel, Steven van de Walle and more. Your academic research and practical insights have been invaluable in bridging the gap between the rhetoric and the reality of performance based reforms in government, an area that often still seems to be dominated by theoretical assumptions and political wishes.

Maarten de Jong, Rabat, October 2015

CONTENTS

INTRODUCTION

Giving money and power to government is like giving whiskey and car keys to teenage boys
U.S. satirist and author and Patrick Jake O'Rourke

The problem with government
Government's reputation to efficiently and effectively perform its tasks seems to always have been controversial at best. Many explanations have been provided as to why governments would be prone to wasting resources and be ineffective. Some of the popular ones are: the absence of market competition for survival, an abundance of easy accessible resources, and the inertia and bureaucratic nature of its institutions. (e.g. Olson 1973: 355-357). Other explanations, more mildly, refer to the nature of government tasks themselves. Providing public goods by definition means that there is no private party willing to buy or supply them voluntarily. This seriously limits comparison of cost prices and quality levels. Also the diversity of interests represented by democratic government, often rules out optimizing for financial efficiency (e.g. Kraan 2011: 6-7)

Whatever the cause may be, the quest for solutions for a better performing government has been unremitting. Especially during the 1990s and early 2000s the New Public Management (NPM) agenda introduced many recipes that were meant to transform the public sector towards more result orientation and efficiency. Although its roots trace back further than NPM, performance budgeting (PB) gained worldwide popularity in the 1990s as part of the NPM reform agenda and is the subject of this research (OECD 2007: 11).

Bringing performance data into the allocation debate
Although it wasn't until the 1990s that Performance Budgeting ideas piggybacked on New Public Management reform agendas worldwide, the ideas themselves were not that new (Redburn et al 2008: 7-12, Joyce 2003: 24). The optimal allocation of resources by government has been one of the most vital questions to all involved in politics and public service as well as to those in the scientific community studying the fields of public administration and economics. To a politician the right allocation may be an allocation representative of his or her ideological convictions or an allocation that rewards his or her constituency or large groups of voters in general. To those involved in public service this may be an allocation that rewards their specific organization or line of work or is in accordance with their professional assessment of prioritization. To those in the scientific community the right allocation can be one that maximizes utility according to an economic model or an allocation that has the largest support base in civil society.

Performance budgeting has been another attempt to help solve the puzzle of allocating public resources in a more objective, systematic way. The notion that the quality of public goods and services should be a major factor in allocating public resources seems to have

slowly conquered the minds of all the aforementioned groups during the last decades (Breul& Moravitz 2007:2). Within government, promoting performance measures as an objective standard for allocating resources has in particular gained popularity with those tasked with budgeting within government. Understandably budget staff will welcome a tool that enables them to distinguish between rewarding a powerful budget claim from just rewarding a powerful claimant (Posner 2009: 18). Politicians have also discovered the appeal of PB as a way of promising more value for money to taxpayers or as a way to curtail public spending to find funds for their priorities. An increasing pressure to be accountable for results has been felt to a more or lesser degree for all those working for the public sector since.

The results of reform

There is a gap in literature when it comes to the studying the impact of performance budgeting reforms (OECD 2007: 58). An answer to the question whether PB has been successful as a reform or not cannot be answered without distinguishing between the different expectations that have been attributed to PB. Broadly, the two main reform objectives for governments around the world have been:
1) To increase transparency of government spending and results
2) A more effective allocation and management of resources

Although PB initiatives have resulted in more transparency (Curristine 2005: 216/17, GAO 2004b: 11-12, Van Nispen & Posseth 2006: 57), its effects on budgetary allocation by parliaments have generally been limited (Frisco and Stalebrink 2008: 11, De Vries& Bestebreur 2010: 237, OECD 2007: 66-67, De Jong et al 2013: 14-15). Operational efficiency gains and internal results orientation are effects that take place within government agencies and are therefore less visible and harder to substantiate. At the same time, several authors have stressed that the benefits of PB reforms are most likely to occur at the agency level (Joyce 2003: 61, Posner 2009: 7-8, Van Dooren 2011: 429). The evidence on the effects of PB at the agency level is mostly anecdotal and often conflicting. That is why this research tries to determine whether PB delivered on some of its promises in managing efficiency in principal-agent relationships.

The purpose of this study

Despite widespread criticism of and disappointment with New Public Management reforms, PB is credited with contributing to improved efficiency and effectiveness in several public sector agencies. The general consensus seems nowadays seems to be that PB can only deliver results if certain other crucial management conditions are met or as Allan Schick already put it over a decade ago: *governments that don't manage for results will not budget for results'* (Schick 2003: 102). The fact that an agency successfully uses performance data to realize improvements or efficiency gains may however be the result of a diversity of factors. Corporate culture, public sector motivation, leadership commitment and discretionary

power of managers have already been identified as important in this respect (Mahler 1997: 538 , Dull 2009a: 273, Moynihan&Pandey 2010: 14, Moynihan&Lavertu 2012: 599-600). Moreover, the causal relationship between successful PB adoption and purposeful PI use may in fact be reversed. In other words, did PB indeed modify public organizations and their steering relationships? Or did it merely codify existing behavior in just a few cases? If the latter is true, the potential of PB to change agency behavior should be seriously questioned. The general research question posed at the start of this research is therefore: *Are result orientation of a government agency and operational efficiency gains achieved through PB?* After taking a closer look at the characteristics of PB success in its most likely form, this question is operationalized into the central research question:

How do underlying cultural and historical factors explain successful PB in government agencies?

In an attempt to further explore the 'DNA' of agencies with a reputation of successful performance management, this study turns to alternative explanations provided by neo institutionalism. More specifically, a number of indicators concerning deeply rooted organizational traditions, the impact of events, reflective openness and cognitive frames were explored as explanations for result oriented behavior. After finishing the case studies, a conclusion is drawn on whether the use of PI was generally supported by coincidental historical and cultural circumstances in these cases. If so, this may provide a credible alternative explanation for the adoption of PB methodology thus weakening the universal claim that PB changes the way a public sector agency conducts its business. However, if the cases do not result in a convincing alternative explanation, the hypothesis that adopting PB does indeed alter behavior of agencies still stands.

Relevance of this research
Internationally, despite plenty of disappointing experiences, PB is widely advocated by institutes such as the IMF, World Bank and OECD as part of the core set of modern Public Finance Management (PFM) reforms along with Multi-Annual Frameworks (MTEFs) and Spending Reviews. Benefiting from recent research, the debate in these circles is increasingly turning to sequencing of reforms thereby acknowledging the necessity of accomplishing a certain base level as a foundation for modern PFM architecture. As a consultant, researcher and lecturer for amongst others the World Bank and the OECD, the author is a participant in this debate. Today, almost all experts in the field stress how non-technical factors heavily influence the design and the success of PFM reforms (Diamond 2011: 1). With regard to PB, the dominance of culture and traditions is increasingly acknowledged as a more important factor in agency management than budget allocation (Schick 2014: 20). The next step for increasing the impact of technical assistance in PFM reforms will therefore require the difficult task of identifying the relevant institutional factors for success and incorporating

these in approaches to reform. This includes explicit recognition of the variation that exists between, but also within national systems.

In recent years, an increasing number of quantitative studies appeared that linked PI use in the public sector to a diversity of factors adding up to over 30 by 2012 (Kroll 2015: 470-471). As mentioned these include factors such as corporate culture, public sector motivation, leadership commitment, discretionary power of managers. Quantitative analysis of data sets can result in important clues on favorable characteristics and circumstances for PI use by public organizations but also has some obvious limitations. Working with a large N-size, inevitably requires simplifying complex variables in order to make them manageable. By relying on abstractions such as pre-determined cultural categories so they can be measured in multiple organizations, no attention is given to specific characteristics and nuances of culture nor to tracing these characteristics to actual PI use. Similarly, another limitation of quantitative work in this area is that it offers little historical perspective on the development of an organization and its PB efforts. Instead, the available cross-sectional analyses and experimental designs have a very short gap between treatment and observed response. For these reasons, qualitative case studies can have added value in this field of research. Especially with complex and intangible factors like organizational culture and leadership, qualitative case study analysis such as this research, are a valuable if not necessary addition to make sense of some of the identified factors resulting in purposeful PI use. This research should be viewed as a step, (albeit a modest one) to bridge this empirical gap by offering a more specific narrative of the contribution of some of these factors.

CHAPTER 1 THE UNFULFILLED PROMISE OF PB – A PROBLEM ANALYSIS

It was Winston Churchill who allegedly said that success is the ability to go from failure to failure without losing enthusiasm. Measured by this standard the performance budgeting (PB) movement should be credited with quite a successful history. Attempts of governments worldwide to structurally integrate their budgets with performance planning and reporting has, so far, not brought all the improvements that were envisioned by many reformers. At the same time the promise of objective performance informed decisions by public sector managers and politicians and occasional success stories have kept these reforms on government's agenda's worldwide for at least the last two decades. A long standing claim made by PB reformists has been that it can contribute to more efficient allocation of public spending and provide tools to cut wasteful ineffective spending. So far there has not been much evidence supporting major reallocation of spending as a result of PB reforms. Other than a lack of visible reallocation decisions, it remains to be seen whether PB reforms did deliver on some of the other promises they held. Results of PB in other areas such as increasing government transparency and agency management look somewhat more promising. This chapter explores the question what is meant by PB, looks at likely implementation difficulties and takes stock of worldwide experiences: disappointments as well as claims to success. The will lead to the central question for this research at the end of the chapter.

1.1   What is meant by performance budgeting?

A multitude of promises and expectations was communicated by performance budgeting reformers at the start of reforms. These can be generally classified into two broad aims. One is to increase transparency of government spending and associated results in order to give voters and the legislative branch better opportunities for accountability and oversight. This requires integrating of performance data into budget documents and requires measuring and reporting of performance information. The other broad aim of PB reforms is to increase effective allocation and management of resources. This not only requires measuring and reporting of PI but also actually using this information to inform decisions made in the budgetary process. Both broad aims of PB are expected to contribute to solving the puzzle of allocating public funds in their own way. A transparent budget proposal is seen as a necessary precondition for making better informed choices. The use of performance data in the debate about different allocation options is expected to result in decisions that fund those programs and policy instruments that have shown adequate results while improving or saving on less effective ones. Savings generated by reducing the funding level of particular programs whilst retaining or increasing spending on performing programs are referred to as improvements in *allocative efficiency.* This can be expected to occur during budget resolution by parliament (OECD 2007: 66-67)

Savings generated by changes in funding levels within a program are referred to as improvements in *operational efficiency.* Operational efficiency improvement as a result of PB

5

happens at the agency level in the budget preparation and execution phases and is therefore less visible than changes within authorized programs (OECD 2007: 66-67).

Although not explicitly envisioned by many PB reformers, another reoccurring claim to success of PB is an increase in result orientation within government organizations. These have to do with providing a greater emphasis on tangible results (including setting of objectives, monitoring performance, planning) by government and improved transparency of planning and reporting documents for their users (GAO 2004b: 11-15, IOFEZ 2004: 41-42 , OECD 2007: 59-61). This type of success is hard to identify and substantiate for non-organization members or distant stakeholders and is equally difficult to define precisely. In an attempt to do so anyway, result orientation by a public sector agency in this study will refer to the degree in which professionals, staff and management integrate information on strategic goals and measurable objectives in their daily actions and decisions.

Although often more of a 'bycatch' than a stated intention of PB reforms, increased result orientation of public sector organizations and its members can be counted as a separate type of result. Figure 1.1 shows a classification of potential contributions resulting from PB reforms. It can be argued that all of these contributions are to some degree interlinked.

Figure 1.1   Potential contributions of PB reforms (De Jong et al 3013:6)
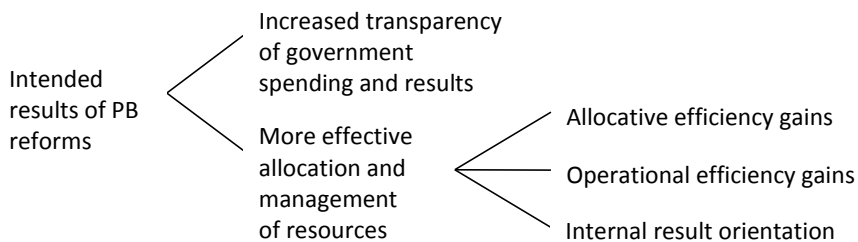
Measuring, reporting and presenting the results of government programs, measures or activities is the key to PB achieving any of these aims. By communicating clear policy goals and targets and monitoring them during and after execution (much like the way this is done with financial information) it becomes possible to base future decisions on past measured performance and draw explicit lessons. Using the budget process as a vehicle, the information on policy results is presented simultaneously with the funding decisions thus making sure that the lessons can be learned during allocation decisions.

Turning to PB, there are several ways of categorizing performance budgeting systems. A common way is to do so according to the degree to which PI is linked to funding. Allen Schick distinguished between two polar versions of performance budgeting using a broad and a strict definition:

Broadly defined, a performance budget is any budget that presents information on what agencies have done or expect to do with the money provided to them. Strictly defined, a performance budget is only a budget that explicitly links each increment in resources to an increment in outputs or other results. The broad concept views budgeting in presentational terms, the strict version views it in terms of allocations. Many governments satisfy the broad definition, few satisfy the strict definition. (Schick, 2003: 101)

Teresa Curristine used a three-fold typology to describe performance budgets in OECD countries (OECD 2007: 21, Congiano et al 2013:229). In some cases, a direct link between performance, resource allocation and accountability is in place (direct/formula performance budgeting). More often, the link is indirect, and planned performance targets and results are used for planning and accountability purposes only (performance-informed budgeting, or PIB). Finally, there are performance budgeting systems that have no link between performance and funding and that use PI for accountability only (presentational performance budgeting). It is not easy to identify a clear relationship between the form of PB system and the emphasis of its intended results. It is clear however that presentational PB is more likely to directly target transparency whereas direct performance budgeting is explicitly geared towards reallocation.

Arguably, a fourth type might be described as a managerial performance approach which focuses on managerial impacts and changes in organizational behavior but may de-emphasize a strong budget linkage (OECD/Von Trapp 2014:2). It should be noted that the line between performance based budgeting and performance management is becoming increasingly blurred. In fact PB today is seen more and more as a subset of performance management rather than a separate process (Schick 2014: 3). Integrating performance into budgetary consequences in some form and at some level seems to remain as the main distinguishing feature of PB.

Notwithstanding the diversity in PB definitions that emerged and transformed over time, using a PB system is commonly associated with the following activities (based on Schick 2003, Robinson&Brumby 2005, OECD 2007, OECD 2015 and Schick 2014):
- setting measurable objectives and performance indicators for government programs
- presenting expected results alongside spending levels in budget documents
- measurement and reporting of results during or afterwards program execution
- evaluation of results and use of this information for strategic planning and budgeting

The type of information that is actively generated, collected and disseminated as part of a formalized performance measurement and reporting system, can be contrasted with non-routine PI such as ad-hoc feedback that is passively received (Kroll 2013: 265). This is an important notion when critically assessing the added value of adoption of PB systems and its

effects on the performance dialogue within and between organizations. As can be noted, this interpretation is broader that some of the older PB definitions in the sense that it not limited to direct use of PI for budgetary purposes and therefore does not distinguish sharply between performance management (PM) and performance budgeting. The problematic separation between PB and PM and has increasing been recognized by the PB community during the last decade or so (Schick 2014: 3).

Insofar goal setting, performance measurement, analysis and reporting are part of the same formalized cyclical routines as the budgetary cycle, the term PB is being applied (as may PM at the same time). This is justified by the expectation that gains in efficiency or effectiveness resulting from such a system will, at some point find their way in the budgetary cycle at a certain level. The effect assumed by reformers is that reforms will lead to what has been labeled as a purposeful PI use, that is, use that is likely to result in efficiency and effectiveness gains (Moynihan&Lavertu 2012: 594). This could for example be the case where PI is used for intra program reallocation or rewarding or dismissing claims during budget preparation.

Governments with an advanced PB system have found ways to translate political priorities into policy objectives that are translated into programs and activities. At one level these are associated with the budget plans as presented to parliament for authorization. This chain of accountability is typical to principal-agent relationships. The agent's internal system of performance planning, monitoring and evaluation therefore usually reflects that of the principal as the agent is subject to reporting and oversight by the principal on the same policy goals. The relationship between strategic performance planning and budgeting as envisioned by most experts in the field PB is illustrated in Figure 1.2.

Figure 1.2   Strategic planning and budgeting cycle (source Redburn 2008: 13)
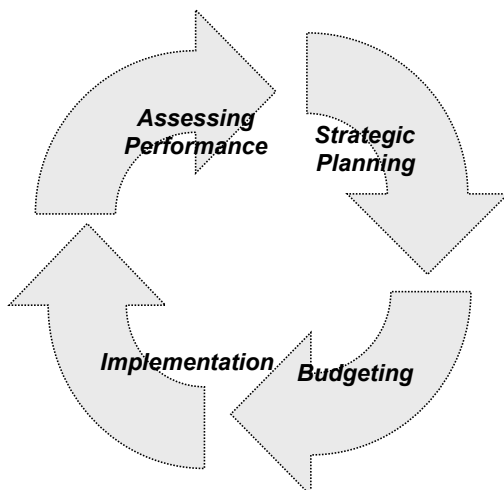
PI is intended to play a role in all phases for planning, target setting, monitoring and evaluation respectively. Arguably the link between assessing performance and the subsequent phases of strategic planning and budgeting are the ones that have proven to be the most problematic as the latter two processes have their own politically driven dynamics in which perspective has to compete with other, often conflicting, considerations such as stakeholder interests and ideological convictions.

As such, PB can be contrasted to the more 'traditional' way the budget process takes place in government. This was articulated well by Aaron Wildavsky when describing the various sources of information that agency staff takes into account when deciding 'how much to ask for' during budget preparation (Wildavsky 1992:90-94):

- Estimates of what 'will go'. As the agency is usually faced with more demand for funding than is available it will asking just below what will be considered 'too much' by the other players in the budgetary process. This way it remains credible when requesting funds for priorities in the future.
- Informal reports on how its programs were regarded, especially last year's
- The letter from the budget bureau, which usually has some statement on how closely this year's budget should resemble last year's.
- The interest of specialized publics in particular programs. Periodic reports from the field on the demand for services may act as an indicator. When the agency begins to notice connections between the activities of supporting interests and calls from parliament, it has a pretty good idea of the support for a program.
- Interest of (influential) politicians in particular programs, mainly 'likes' and 'dislikes' of influential politicians.

Perhaps not surprisingly none of these sources explicitly refers to performance indicators nor to any systematic form of assessment of effectiveness and efficiency like evaluation or spending reviews. In reality any government agency can be expected to demonstrate 'traditional' budgeting behaviour to some extent as PI will never be the sole determinant of budgetary decisions in the public sector.

1.2   PB as part of the New Public Management legacy

Especially during the 1990s and early 2000s the New Public Management (NPM) agenda introduced many recipes that were meant to transform the public sector towards more result orientation and efficiency. Although its roots trace back further than NPM, performance budgeting (PB) gained worldwide popularity as part of the NPM reform agenda and is the subject of this research.

This does not mean that the concept of PB is all that new. Records of the U.S. Federal government's interest in performance budgeting go back to the early 20th century following

recommendations of the Taft Commission in 1912 (Redburn et al 2008: 7-8). Other U.S. initiatives in the previous century included chronologically: The Hoover Commission (1949), Budget and Accounting Procedures Act (1950), DoD's Planning, Programming, and Budgeting System (PPBS) in the mid-1960s, Management by Objectives (MBO) in 1973 and Zero Based Budgeting in 1977. The series of subsequent initiatives over the years inspired Allan Schick to cynically note that: 'performance budgeting is an old idea with a disappointing past and an uncertain future' (Schick 2003: 100). Nonetheless the question remains whether performance budgeting reforms so far have had a serious impact on the way government does business and if so what impact exactly.

When PB gained worldwide popularity, thinking about public administration was dominated by New Public Management (NPM). It can however be argued that government performance and government efficiency was politically identified as a major problem in the period preceding NPM. In this period the roots for the broad popular appeal of PB may have largely been laid. NPM can be viewed as one of the three periods of public administration reforms since the 1980s in OECD countries (OECD 2105: 17-19):
   a. Receding government (1980-1990),
   b. New Public Management (1990- 2005), and
   c. Basic government (The current period from 2005).

The notion that the lack of market incentives in the public sector lead to enormous waste of taxpayer's money with no accountability for results, was starting to get used successfully by politicians. In the 2015 OECD report on Value for Money in government, this period is described as follows:

The period of receding government is commonly associated with the Reagan administration in the US and the Thatcher Cabinet in the UK, but there were many governments in other OECD countries that subscribed to similar ideas and programs. Much emphasis was put on the reduction of government tasks, particularly the tasks of central government by deregulation, privatization, decentralization, expenditure reduction and fiscal alleviation. The idea was that private sector growth and prosperity was hampered by too high tax and expenditure levels. A popular saying of the period was that "government is the problem, not the solution". Theoretical underpinnings of these ideas were partly found in supply side economics and partly in public choice economics. Supply side economics showed that essential incentives for growth in the private sector could be impaired by too high levels of taxation, public choice theory showed that without strong constitutional and legal constraints, the incentives motivating bureaucratic and political behavior, would lead to ever expanding public sector that would eventually overwhelm the private sector of the economy (OECD 2015:17)

As noted in the above quote, Margaret Thatcher in the UK and Ronald Reagan in the US were notable pioneers in verbalizing criticism on the functioning of government. This is illustrated by a number of quotes from former U.S. President Reagan in Textbox 1.1.

Textbox 1.1   Quotes of President Reagan on government's effectiveness and efficiency (source Adler&Adler 1998)

'Government is like a baby. An alimentary canal with a big appetite at one end and no sense of responsibility at the other'

'No government ever voluntarily reduces itself in size. Government programs, once launched, never disappear. Actually, a government bureau is the nearest thing to eternal life we'll ever see on this earth'

'Government's view of the economy could be summed up in a few short phrases: If it moves, tax it. If it keeps moving, regulate it. And if it stops moving, subsidize it'

'Government always finds a need for whatever money it gets'

'Government does not solve problems; it subsidizes them'

'The best minds are not in government. If any were, business would steal them away'

'The most terrifying words in the English language are: I'm from the government and I'm here to help'

It is important to note that the rhetoric and powerful sound bites about the way government is managing public funds has by no means lost their appeal throughout the years. Fueled by the popularity of this view of government and possibly the first wave of reforms, policies to shrink government, reduce its impact on society and increase accountability to taxpayers continued in the 1990s and the first decade of this century. In the UK, some claim that the subsequent governments that superseded Thatcher's were in fact 'on a Thatcherite autopilot' (Jenkins 2006: 334). Textbox 1.2 shows quotes of the current and previous U.S. President on managing public spending with referral to reallocation using PB methods.

Textbox 1.2   U.S. Presidential quotes on the use of PB methods to manage public spending

> *There comes a time when every program must be judged either a success or a failure. Where we find success, we should repeat it, share it, and make it the standard. And where we find failure, we must call it by its name. Government action that fails in its purpose must be reformed or ended.*
> **Presidential candidate George W. Bush in a campaign speech in Philadelphia in 2000**
>
> *We committed to change the way our government in Washington does business so that we're no longer squandering billions of tax dollars on programs that have outlived their usefulness or exist solely because of the power of a lobbyist or interest group. We can no longer afford to sustain the old ways when we know there are new and more efficient ways of getting the job done. Even in good times, Washington can't afford to continue these bad practices. In bad times, it's absolutely imperative that Washington stop them and restore confidence that our government is on the side of taxpayers and everyday Americans…..Our problem is not just a deficit of dollars. It's a deficit of accountability, a deficit of trust.*
> **President Obama January 7th 2009 when appointing his first Chief Performance Officer**

The period of receding government was followed by New Public Management. The many studies that describe the contents of New Public Management share a number of central elements. Other than PB these elements comprise (OECD 2015:18):

- Separation of policy execution from policy development;
- Stimulating competition among service providers, by allowing private suppliers to provide collectively funded services and through demand financing (consumer subsidies and vouchers);
- Loosening of standards of operational management both in policy development and policy execution ("let managers manage");
- Financing executive agencies on the basis of output targets;
- Outsourcing of intermediate production for both core ministries and executive agencies to the market.

Theoretically, NPM is inspired by amongst others public choice and principal-agent theory (Van Kersbergen&Van Waarden 2004: 148). The NPM ideas were proposed and promoted by governmental commissions, task forces, public administration academics as well as private consultants. Many OECD nations introduced forms of PB in this period. New Zealand pioneered budgetary reforms in the 1980s and 1990s that in many ways have not been repeated since in terms of ambition and bluntness. In the US, Vice President Gore chaired the National Performance Review that was established in 1993 in an attempt to apply the ideas put forward in the book 'Reinventing Government', the classical NPM work by David Osborne and Ted Gaebler (Redburn et al 2008: 10-11). The U.S. Federal Government Performance and Results Act (GPRA) of 1993 drew on the lessons learned from earlier

experience. Under GPRA, whose requirements took effect in FY 1999, agencies are required to issue performance plans and reports that are loosely linked to the budget. The Netherlands joined the performance budgeting community at the turn of the century. From 1999 onwards, the traditional structure of the budget was extensively modified into a performance budget under the acronym VBTB, translated as Policy Budgets and Policy Accountability.

Many believers in the NPM ideas in general and PB in particular have since developed into their biggest skeptics as they became disappointed with the results of these reforms. In addition, an increasing amount of academic studies provided evidence of the limited or even counterproductive effects of NPM motivated measures. Confusingly, there also seems to be a broad consensus among practitioners that today's governments, after the (partly unsuccessful) implementation of NPM ideas, are generally better off compared to the way government was run prior to these reforms. This may be why international organizations such as the OECD, World Bank and IMF remain convinced that NPM type of reforms are at the core of their advice to the nations they assist. It is not yet clear what the post NPM era will look like exactly, but despite the universal appeal of many NPM ideas, it looks like NPM's heyday is well behind us. In the post NPM public sector landscape however, many traces of NPM reforms can be found.

One powerful idea in particular has been the separation of policymaking and policy execution as government was supposed to 'steer and not row'. This has resulted in performance based (quasi-) contracting and monitoring arrangement with (semi) autonomous government agencies (e.g. Rhodes1997: 48-49). The introduction of PB supplemented the separation between policy design and policy execution introduced simultaneously in many countries. By introducing (quasi-) contracts that specified the agent's production levels, performance standards and prices, NPM reforms aimed at giving the principal a better position to align resources and agency activities with its policy objectives and exert control over policy execution by agencies. This way, assumptions from principal agent theory have had significant impact on shaping the post NPM public sector landscape in many countries (Ter Bogt 1997: 53-55, Bell 2002: 11). Additionally NPM assumes that by substituting input controls for output and outcomes, like the PB recipe prescribes, managers gain discretionary power to achieve the results they are held accountable for.

The current period from around 2005 is characterized by revising, adjusting or sometimes even abolishing some NPM reforms. Although still early to tell, a new trend may be an increased focus on front office tasks at the expense of back office activities. This may include separating financing of agencies from output or outcome controls (Kraan 2011: 5, OECD 2015: 309). The current economic and fiscal crisis is likely to overshadow any attempt to introduce new large scale reforms in the foreseeable future. As governments will find

themselves increasingly strapped for resources, the actual saving potential of the past reforms and lessons learned from these reforms may gain renewed interest.

As for PB, its popularity may have reached its peak some years ago. Those countries who adopted it are continuing to struggle with ways to make PB work or improve on its results. Examples of this are the introduction of revisions of the earlier systems in which performance is increasingly decoupled from the budget process, such as in the US (GPRAMA 2010 – Moynihan&Kroll 2015: 6-16) and the Netherlands (Accountable Budgeting from 2012 – De Jong et al 2013: 19-28). A set of instruments gaining (renewed) popularity in recent years are performance based policy evaluations. Shaped to result in relevant options for budgetary decision making, performance based evaluations have been emerging in countries such as Chile, Canada, the Netherlands and South Korea (Robinson 2014:12-13, Van Nispen 2015 forthcoming 2015). The latter adopted a version of the former U.S. Program Assessment Rating Tool, which despite lacking budgetary impact in the US, has been used with some success by the Koreans since its introduction in 2005 (Kim 2010: 183-4). In some cases these experiences build upon a long tradition with conducting spending reviews such as in the Netherlands (Schoch & Den Broeder 2013: 15) and Canada (Robinson 2014:21). Policy evaluation and spending reviews can be placed in the broad PB tradition as it attempts to combine funding and results (Schick 2014: 18). In addition the increased availability of information on performance allowed for more evidence based ex-post policy evaluations.

1.3   Challenging PB theory
In spite of PB's logic appeal and its rapid expansion, anecdotal evidence from practitioners as well as several disciplines of social science offer a multitude of explanations why the predicted benefits of PB doctrine do and will not materialize. Much of this criticism also applies more broadly to systematic evaluation of policy effectiveness. Many assert the view of public policy as being sufficiently predictable and measurable to allow for the kind of systematic or even mechanical analysis that PB seems to assume. For this reason PB, according to the traditional strict definition (that is, using performance information in the annual budgetary context), meets even more criticism than the contemporary more broad definition of PB that also accommodates performance evaluation and performance management. Nonetheless even these forms of performance information face some criticism for resting on shaky assumptions.

To agency managers, who are usually busy battling unexpected crises that can only be cured by resources rather than strategic thinking, performance management is little more than a distraction (Moynihan 2008: 16). Faced with the complexity of real world policy dilemmas, bureaucrats understandably opt for other ways to select policy alternatives than comprehensive systematic analysis. It can be argued that their scientifically 'flawed' methods are often superior to any futile attempt at superhuman comprehensiveness (Lindblom 1959: 88). In this light, coupling complex and outlandish activities such as strategy

and policy analysis to the budget process may prove just too ambitious. In a critical appraisal of PPBS (one of PB's important predecessors in the 1960s) Aaron Wildavsky stated that the shotgun marriage between policy analysis and budgeting should be annulled because policy analysis is hard enough to accomplish without having to meet arbitrary and fixed deadlines imposed by the budget process (Wildavsky 1969: 196).

Neither should it be surprising that the political environment in which budget allocation takes place, does far from guarantee a predictable use of the PI that is generated. It would take a totalitarian regime to fully embrace a normative theory of budgeting, for this would imply the end of politics (Wildavsky 1992: 429). In order to be able to deal with unpredictability and complexity, political decision makers (and their staff) often prefer loosely formulating goals, redundancy in information sources and non-cyclical decision making (In 't Veld 1999: 76-82). The view that performance measurement is too simple an approach for the diversity and complexity in the public sector was expressed by Beryl Radin as she confronted six assumptions that constituted what she called the 'unreal and naïve approach' of the performance movement, with reality (Radin 2006:207):

- Information is already available
- Information is neutral
- We know what we are measuring
- We can define cause-effect relationships in programs
- Baseline information is available
- Almost all activities can be measured and quantified

In summary, the assumption of linear cause and effect relations, as well as clear goals and planned change does not survive the reality check (Van Dooren 2011: 427). Apart from overcoming numerous technical challenges, studies attempting to explain a lack of success PB reforms have had so far, reveal that success is determined by some major factors that, as far as we know, remain largely unaffected by the reforms themselves. Many factors may affect decisions to integrate the results of performance assessment into a strategy and budget proposal by a government agency.

In order to be capable to learn from performance assessments to improve effectiveness and efficiency, it is assumed that this type of learning is supported by well-functioning policy and management systems. These systems include: governance and accountability arrangements, political decision making, human resources, policy analysis and advice, performance management, budgeting, monitoring and reporting and evaluation. In many government organizations it remains to be seen whether these conditions are being met (Scott 2003: 57). The notion that PB can only deliver results if certain other crucial management conditions are met was articulated by Allan Schick as: *governments that don't manage for results will not budget for results'* (Schick 2003: 102) . But even if PB is accompanied by a set of tools such as balanced scorecards, internal contracts systems or even pay-for-performance, this

does not mean that these tools indeed play a serious role in decision making because the relational aspects are often valued more highly than the formal aspect (Schick 2003: 91).

A rather obvious but nonetheless still often overlooked condition is formed by organizational culture. In government organizations a focus on compliance instead of performance can provide an additional obstacle result orientation and therefore learning behavior (Schick 2003: 86, Osborne & Plastrik 1997: 258). Generally, it seems that knowledge about result orientation and learning behavior of organizations was ignored when designing the new formal performance budgeting and performance management structures (Moynihan 2005: 203). In a given government agency, displaying result oriented behavior by conducting assessing or debating effectiveness may be supported or punished by the organization's leadership and culture. Cultural obstacles can prevent managers from using PI in a meaningful way. Examples are a lack of authority or interest in change and fear of 'rocking the boat' (Hatry et al 2003: 13). The cultural circumstances necessary for PB to work will be explored deeper in CH.4.

Even if technical challenges are overcome and PB produces relevant information that is picked up within public organizations to learn and improve, the political environment in which they operate may not. The span of attention of any parliament is too limited to oversee all areas of public policy with the same level of scrutiny. Most policy decisions, including budget allocation, are therefore debated within the limited community of policy specialists, parliamentary commissions and political leadership. For these and other reasons, many have argued that policy process is incremental by nature, at least in a democratically run government (Lindblom 1959: 84-85, Wildavsky 1961:184-185, Wildavsky 1992: 436). This observation is easily confirmed by most practitioners. Others assert this view by claiming that longtime periods of calm, incremental policy adjustment are altered by short intervals of turbulence when a policy theory is subject to revising in the public arena (True et al. 1999: 176-180).

For political leaders who are committed to performance based decisions, the PI informed, specialist perspective competes with other perspectives that put issues on the political agenda such as interest groups, political party affiliation and public opinion (Posner 2004: 8-14 , Conlan et al 2002: 4, 12-13). Formally, the agreed upon ideas and policy ideals might have been 'frozen' in principal-agent relationships and internalized in an organization's management control and performance management systems. However one ought to acknowledge that this might only reflect a small portion of the relevant political rationality. Political pressure can not only lead to obstruction of gathering reliable and objective information by bureaucrats (Van Nispen 1993: 132-141), but can also result in the use of PI for cynical purposes where cherry-picking of successes replaces objective reporting (Diamond 2005: 15).

Unexpected results often associated with performance measurement and reporting are perverse effects, especially when financial rewards are directly at stake. In these cases the organizations complied with performance targets in ways that actually damaged the interest of their clients. This is illustrated by a multitude of cases worldwide that vary from selective presentation of results to data manipulation and fraught. A recent example was provided by the U.S. Department for Veteran Affairs, where a scandal was uncovered in 2014 that included data manipulation. In this painful case records were manipulated to keep patients off the official waiting list to meet performance targets of timely treatment. This resulted in dozens of veterans dying while waiting for medical treatment. Ironically a comparable case of manipulation of hospital waiting lists was uncovered over a decade earlier regarding performance targets of the British National Health Service (Smith 1995: 90-92). Another common unexpected result of PB goal displacement, meaning that measurable targets get a disproportionate amount of attention and resources. This gives quite another meaning to Peter Drucker's[1] popular NPM mantra '*What gets measured gets done'.*

Regarding the relevance of PI use for budgeting, once it finds its way in critical performance assessments, there is some additional reason for discouragement. According to the 'Law of Policy Accumulation formulated by In't Veld, policy makers tend to respond in the same way regardless an instrument's effectiveness or ineffectiveness, namely by calling for more of the same (In 't Veld 1998: 153-161). In budgetary terms this means more funding for the same policy instruments even if they appear flawed.

The previous paragraphs included many possible misconceptions underlying PB, and other reasons why reforms would be prone to failure. Paradoxically some of these reasons exactly reflect the reasons why, proponents claim, PB should be adopted in the first place. Summarizing the case against PB, four arguments appear as being central:
- Politicians understandably prefer to express policies in loosely formulated goals
- Policy success is too difficult to measure by a comprehensive set of indicators
- Politicians and public sector managers are not interested in performance data
- The implications that measured performance should have on funding levels remain unclear

1.4   Assessing PB results

Given the rather fundamental critique on PB in the previous section, it may not be surprising that, despite its intuitive appeal and widespread adoption, putting performance budgeting it into practice has proved challenging. While the budget provides a unique crosscutting mechanism to collect performance information, Schick notes that with few exceptions, performance budgeting has not become the government's budget process (Schick 2014: 9) Rather, for most countries it is an accessory to the budget, adorning spending decisions but

---

[1] Peter Drucker was a popular management author in the 1990s and was much quoted by NPM reformers during that decade.

not fundamentally changing the way they are made (OECD/Von Trapp 2014: 3). More specifically, experience in OECD countries as well as academic research demonstrates that although PB initiatives have resulted in more transparent budget documentation (Curristine 2005 in Redburn 2008: 216-17, GAO 2004b: 11-12, Van Nispen& Posseth 2006: 57).

When it comes to improving effectiveness and efficiency, the successes of implementing a result oriented budgeting system seem to be absent or at least less visible. There is no clear evidence that supports the claim that performance budgeting has had a direct impact on fiscal discipline or reallocation of funding (OECD 2007: 67). As policital priorities not agency performance, drive macro budgetary choices (Behn 2003: 590), its effects on budgetary allocation by parliaments have generally been limited (Frisco and Stalebrink 2008:11, De Vries&Bestebreur 2010:237, OECD 2007: 67, De Jong et al 2013: 14-15).

In the Netherlands, after the evaluation of VBTB, not the budget structure but policy evaluation was presented as the main source for performance improvement. In New Zealand, some of the much-praised pioneering reforms of the 1980s and 1990s have been quietly modified or even reversed during the last decade (Chapman& Duncan 2007:21). The New Zealand Auditor General Kevin Brady reported that the use of PI by internal users to improve public service effectiveness has been disappointing (Brady 2008: 17). With regard to the U.S. Program Assessment Rating Tool, arguably the most ambitious comprehensive effort to link performance and budgeting of recent times (see Text Box 1.3), the results present a mixed picture. When developing funding proposals, it seems that the President's Office of Management and Budget (OMB) did take performance scores into consideration to some extent (Gilmour and Lewis, 2006: 747-748). When looking at program funding as authorized by Congress however, the role of the performance ratings seemed to have evaporated quickly. Moreover the role PART examinations played in debates about funding in Congress seemed to be limited at best (Frisco and Stalebrink 2008:11)[2].

---

[2] A 2008 study by Frisco and Stalebrink indicated that reference to PART by the 109th Congress was minimal. About 6% of congressional hearing reports having a PART related content, 4.74% of 1,033 for the House and 7.79% out of 655 for the Senate. Moreover, only a third of the cases where House members mentioned PART, they did so in the context of specific program results against about half of the cases of recorded use by Senators. Combining these data one can ~~to~~ conclude that performance data generated by PART were considered by Congress about 2.5% of the time.

Textbox 1.3 Explanation of OMB's Performance Assessment Rating Tool (PART,2002-2008)

PART was introduced as an element of the more elaborate Presidents' Management Agenda and within 5 years the effectiveness of 973 federal programs, representing 96% of total budget expenditure was assessed as being either 'effective', 'moderately effective', 'adequate', and 'inadequate' or 'results not demonstrated'. The grade a program received by the OMB was based on the answers to a questionnaire that was standardized for seven different categories of programs. The program's grade was one of the factors contributing to the proposed funding level of a program in next year's budget, alongside political priority and the overall fiscal outlook. The results of PART reviews were posted on www.expectmore.gov for anyone to consult.

An explanation for the limited success in engaging Congress in the PART process was the fact that Congressional staffs were unlikely to accept conclusions about a program's performance without seeing the detailed evidence to support it, something OMB's summaries do not provide. This was especially true in those cases where OMB's conclusions didn't match their own assumptions (Brass 2006: 49/50)

Although one can have serious doubts about the attempts to 'rationalize' budgetary decision making using PB, some claim that performance management reforms can change managerial behavior and PI does get used, be it at a different place and time. Several authors have stressed that the benefits of PB reforms mostly occur at the agency levels and in the budget preparation and budget execution phase and not in the budget approval phase (Joyce 2003: 36-37, Hammerschmid et al 2013: 5, OECD/Von Trapp 2014: 4). With regard to operational efficiency, occasional success are reported. For example in his testimony to the U.S. Senate Budget Committee on October 29th 2009, Paul Posner asserts:

'The performance test for management reforms is whether they produce sustainable results in decisions and program implementation. From this vantage point, the federal agencies in fact have demonstrated real progress in using PI to manage their ongoing activities. Whether it be framing new ways of thinking about goals, or assessing and overseeing employees, contractors or grantees, performance data have given the agencies powerful new tools to reach their goals. For instance, federal agencies are using outcome data to allocate their own staff to areas needing greatest attention. Grant making agencies are using performance outcomes achieved by grantees as a basis for providing bonus funding rewarding those state and local projects achieving notable results. Leading examples illustrate the payoff from performance management and budgeting for agency effectiveness and accountability (…professor Posner continues by singling out 4 examples of federal agency that made improvements using performance metrics)'

Examples of performance informed innovations by public agencies provide credible testimony for a more positive appraisal of the contribution of PB to efficiency and effectiveness of government.

Less visible and even harder to substantiate are claims made that PB enabled better control of government agencies and installed result oriented behavioral incentives within agencies (GAO 2004a: 10-12, IOFEZ 2004: 42, OECD 2007: 59-61, Posner 2009: 7-8, MinFin 2011: 24-25, Van Dooren 2011: 429). In his testimony, Posner also refers to the reoccurring claim that PB has the potential to increase result orientation within government organizations. This has to do with providing a greater emphasis on tangible results (including setting of objectives, monitoring performance, planning) by government. In some cases it is reported that PB adoption provides a necessary spark for an agency to critically assess its results while in others it seems to remain merely a compliance exercise that lacks real impact. In some cases a formal system of performance measurement and management seems to smoothly complement the way an agency conducts its business while in others it is claimed that such a system erodes professional commitment and ethics and may even hamper agency performance (e.g. Robinson&Brumby 2005: 48). Perhaps surprisingly, some note that the assumption that the NPM-ideal type agency model[3] enhances performance of public sector organizations remains largely untested as the claim of an increase in quality and efficiency of service delivery has been often assumed but seldom well documented (Verhoest et al 2012:4). With regard to using PB for agency management this is even more surprising since PB has been promoted as a solution to smooth if not solve information asymmetry, the central problem in agency theory[4]

An answer to the question 'Has PB worked or not?' can necessarily only be given with a particular aim of this reform in mind. Not surprisingly, given the multitude of promises and expectations associated with PB at the time, only a subtle answer can be provided to this question. Summarizing, the evidence supporting success can be considered promising when it comes to transparency. With regard to efficiency the evidence is generally discouraging (allocative) or anecdotal (operational and result orientation). This is illustrated in Figure 1.3.

---

[3] This NPM-ideal type agency model is considered to be a model based on performance contracting between a principal (e.g. a ministry or agency headquarters) and a public agent (agency or agency unit). The principal exercises performance control and uses PI for budgetary purposes while the agent is managed in a performance driven manner.

[4] Principal-agent theory is concerned with ways to get an agent to behave in the interest of a principal. The central dilemma in principal agent theory is information asymmetry that occurs when the agent has an information advantage over his principal. The agent thus has the possibility to serve its own interests at the expense of those of the principal. According to public choice theorists the agent can be expected to behave this way, leading agencies to behave inefficiently.

Figure 1.3   Aims of PB where evidence regarding effects is scarce and inconclusive

Intended results of PB reforms
- Increased transparency of government spending
- More effective allocation and management of resources
  - Allocative efficiency gains
  - *Operational efficiency gains*
  - *Internal result orientation*
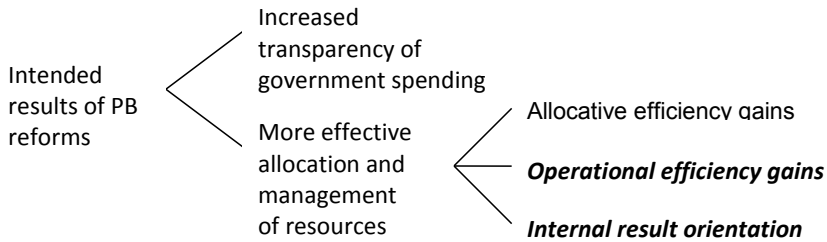
Why PB is here to stay

Regardless of the sparse evidence of agency management and politicians actually integrating PI and their strategic and budgetary decisions, the idea of PB seems powerful enough to keep an irresistible appeal on both politicians and agency leadership. To elected officials PB can have significant symbolic value because the ideas (systematic performance driven policies) and intended results (transparency, efficiency) just sounds too good to disagree with. This does not mean that they will engage in the tough labor of assessing effectiveness and efficiency and engage in the difficult choices that result from this. Political support for PB reforms has therefore more than once been characterized as 'a mile wide and inch deep' (quote Phil Joyce from interview in 2007).

To central budgeting officers and agency budget staff, PB can be a valuable tool because it adds legitimacy to the budgetary process and their power over or within an agency. Moreover, one of the reasons PB is advocated by so many in politics and public sector leadership may be that politicians and budget staff both assume the information could be valuable to the other party (Moynihan 2008: 67). In addition, as noted in Section 1.2, in spite of the disappointment articulated over the effects of NPM reforms, there seems to be little support to decisively roll them back either. Nor did a new comprehensive set of far reaching reforms appear at the horizon to provide an alternative to those seeking to reform government. Given the bumper sticker appeal of PB and institutional factors supporting its promotion, it seems justified to expect the concept of PB to remain dominant on many reform agendas around the world for some time to come. The challenge for practitioners will be to use this reform potential in a way that takes into account the lessons learned from past experience.

The quality of performance assessment and the formal incentive structure have been the traditional focus of practitioners and scholars alike when studying the success of PB attempts in government. Studies attempting to explain the lack of the success reforms have had so far, reveal that, apart from overcoming numerous technical challenges, success is determined by some major factors that, as far as we know, remain largely unaffected by the

reforms themselves and seem to have been largely ignored by the performance based budget community. This will be the further focus of this study.

1.5   Towards a research question

The degree of success of PB reforms can be viewed in terms of general flaws in its design and methodology but can obviously not be viewed entirely separate from differences in implementation. Despite seemingly similar circumstances, at some places at the agency level *de facto* adoption of PB doctrine seems to occur in addition to only *de jure* adoption elsewhere. In other words, there apparently are circumstances at play that determine the success of PB reforms. Whether these circumstances are a common factor among PB successes, let alone if they have any predictive value for de facto PB adoption is currently unknown. The important point to make here is, that the cause-and-effect relation between successfully implementing a formal result oriented budgeting system and an organization's general result orientation and its tendency to use PI, might be confused.

Given the expectation that PB will remain dominant on reform agenda's around the world, it makes sense to look critically at those cases that might fit criteria for success and under which circumstances these occurred. When talking about the conditions for success of performance budgeting, Allen Schick noted in 2003:

'How much influence performance budgeting has depends on who else is at the table-that is, the orientation of the politicians and managers who make budget….Performance budgeting can thrive only when it is embedded in managerial arrangements that make results paramount. This writer is not aware of a single sustained implementation of performance budgeting that was not accompanied and reinforced by transformations in public management that enhanced performance. Governments that do not manage for results do not budget for results, even if they install the outward trappings in performance budgeting…..
….Rather than being the locomotive that drives government to change, performance budgeting is the caboose that confirms the transformations that have been made. To achieve true reform, it may be better to follow the parade than to lead it'. Schick, A. (2003: 102).

If Schick's observation is correct that PB usually fails when not embedded in broader reforms that enhance performance, what then is 'atypical' in those cases where success of PB reform is claimed and can seemingly be substantiated where others failed? Are there circumstances that coincidentally led to this success? More importantly was PB indeed a helpful tool under these circumstances? If so more knowledge about these particular circumstances could help budget departments readdress their requirements concerning PB and attain a more effective approach.

If, to the contrary, the contribution of PB cannot be determined in the cases where success is claimed, it appears we are facing a classical fallacy of circular justification with regard to the success of PB: Attempts to increase consideration of PI by public managers are successful insofar these managers work for data driven organizations with a high result orientation. Instead of *modifying* decision making in the budgetary process, the reforms may have merely *codified* existing practices. In other words: does successful PB implementation **create** favorable circumstances for purposeful PI use? Or can successful PB implementation **only result from** favorable circumstances for purposeful PI use?

When assuming codification as the most plausible explanation for PB success, additional doubts should be cast over the usefulness of investing in PB reforms for the purpose of better budget allocation. Expanding on Schick's observation that governments that don't manage for results will not budget for results, the question seems justified: is there added value in adopting PB reforms for a government organization that already manages for results? If not, why should one invest in these tools at all? After all, we already know that if the right circumstances are absent, PB reforms don't deliver the intended results. Additionally, if the right circumstances are present, the results sought for by PB would be likely to occur regardless of having these reforms. In that case a claim to successful PB would be similar to a claim of having a machine that is able to create sunny weather, but so far has only had encouraging test results in Sicily or Southern California. The general question to start this research is therefore:

Are result orientation of a government agency and operational efficiency gains achieved through PB?

This central question will be refined in the next chapters after taking a closer look at the potential value of PI use form a budgetary point of view. To understand the dynamics of the internal organization, theory on performance management, management control, the learning organization and organizational sources of power will be of assistance. Depending on the case studied, some theoretical knowledge regarding a specific policy field may be required in order to understand the expected relationship between objectives, means and results. Chapter 2 will elaborate further on this question to determine what exactly will be the right focus of this study. Since investigating this questions requires studying alleged cases of PB success, a better definition is required of what exactly constitutes PB success. This requires a further exploration of PB theory and its often implicit assumptions.

CHAPTER 2   BUILDING A MICRO MODEL OF PB

As a wise yet unknown man once said: *A cynical person is an idealist who, at some point, made the mistake of turning his ideals into his expectations.* This seems to apply in particular to many in the performance budgeting community given the enormous expectations and in many cases disappointing results of this reform. But aside from some understandable bitterness, a serious question remains on how serious the problem of disappointing results of performance budgeting (PB) actually is. Did PB fall victim to unrealistic expectations? Have these reforms failed altogether? Did they succeed just partly or temporarily? Or do they just require more patience? In the previous chapter the aims and results of PB reforms were addressed at a macro level. In this chapter PB theory will be further analyzed to more closely define what PB success would look like in its most likely form. Surprisingly this is a rather unexplored activity as noticed by Moynihan:

'Governments have never been so awash in performance data, mostly because they are required to collect and report it. The wealth of performance data contrasts with the poverty of the theoretical and empirical justifications for performance-reporting requirements. We have poor theories of PI use, largely informed by a combination of common sense, some deeply felt assumptions of how government should operate, and a handful of success stories'. (Moynihan 2008: 5).

In this chapter the potential value of producing and disseminating PI is critically appraised by looking at the budget process, the possible role of PI in budgetary and policy decisions and the contribution these decisions may have to efficiency and result orientation. Often neglected in PB theory, the link between measuring and reporting PI and achieving gains in effectiveness and efficiency rests on a number of often unarticulated assumptions that form key elements of PB theory. By studying these potential links and making them explicit, a *micro-model of PB* will be created. This chapter ends with a refined research question and the identification of additional relevant factors that need to be taken into account when answering it.

2.1   The budget process and its actors

Chapter 1 demonstrated that the evidence supporting success of PB can be considered promising when it comes to transparency. With regard to efficiency, the evidence is discouraging (allocative) or scarce and inconclusive (operational efficiency and result orientation). As it was concluded that PB reforms will remain dominant for some time to come, this research aims to address those aims of this reform that are most underexplored (see Figure 1.3). To study PB's possible contribution to generating operational efficiency gains and increasing result orientation in public sector organizations, the processes and relationships vital to PB success are to be addressed. As PB reforms are intended to provide tools for making budgeting decisions, a closer look at the major actors in the budget process,

the budget process itself and the corresponding levels of interaction will be helpful in more closely identifying where to look for success.

Actors

A number of key players can be identified as being relevant in the implementation of PB system in national government systems. Not surprisingly, these same actors are the ones that play a dominant role in the budget process[5]:

- *The Legislature* consisting of Congress or Parliament usually under a system of bicameralism. Within these institutions permanent committees have usually been established to oversee particular policy areas and ministries. Of these the budget committee is of particular relevance as are the committees on which a particular policy field to which a performance budget applies. The Legislature is supported by one or several staff bureaus that have an important role in the selection of information and its presentation to elected officials.

- *Executive Leadership* consisting of the Presidency, the PM or the Council of Ministers. This is the place were conflicting interests within the Executive Power are resolved like for example budget claims or overlapping responsibilities. The Executive Leadership is also supported by one or several staff bureaus. If one of these is tasked with budgeting in particular, like OMB in the U.S, it is treated as a separate actor in the budget process here (see National Budget Office).

- *The Political Principal* can be a minister/secretary or any other political figure with formal responsibility for a particular policy area. His or her office acts as principal to the public sector organization that is tasked with policy design and overseeing policy execution. The Political Principal's office will have a unit that is tasked with budget and performance issues. Therefore this unit maintain relationships with both policy units and the agents that execute their policies.

- *The National Budget Office* consisting of the entity within government responsible for the PB guidelines and regulation. This is usually the Ministry of Finance or a unit for administrative reform directly under the President or PM as is the case with OMB in the US. The Budget Office or its equivalent often acts a second principal for a agency as its regulations need to be met in addition to the objectives of the policy principal whose policy it has to execute.

---

[5] The exact definitions of these actors obviously vary per individual country. The terms here are abstracted based on the majority of OECD nations with especially the federal US and the Netherlands in mind as representatives of two different systems within this frame.

- *The Government Agency (or agency)* can be any public funded organization responsible for executing government policy including policy directorates, (semi) independent agencies or institutions performing public tasks at arm's length of government. The agency acts as a direct agent to Political Principal and often to the National Budget Office as well. Therefore the agency headquarters will maintain a performance dialogue with these entities. Within the agency, units are tasked with the administrative rules of the budget office (budgeting, reporting, performance monitoring) and units tasked with the agency's primary processes associated with execution of the principal's agenda. In its turn the agency can act as an agent to other subordinate agencies or regional branches and be engaged in an internal performance dialogue with them.

  An agency's budget department's role in the budgeting process can be of a somewhat schizophrenic nature as it internally attempts to control spending and protect the interests of its principals while externally it may find itself acting as an advocate for increased spending on its organization. This can be justified by the counter-roles assigned to other players in the budgeting process as part of a system of checks and balances (Wildavsky 1992: 89).

- *An external auditor or evaluator* can be tasked with assessing the performance of programs, sometimes linked to funding levels.

As noted in Section 1.4, PB systems potentially increase the role of the budget office over policy and the involvement of policy units into budgeting. The role of the budget office as described here applies in particular to countries with more advanced PB systems that share a top-down system that is designed for comprehensive coverage of the entire budget. Examples are Canada, the US, the UK and the Netherlands. Top down PB reforms enable a more uniform and systematic approach and result in higher availability of information at the center of government and better coordination and monitoring whereas in bottom up systems individual agencies enjoy a greater degree of freedom in designing and executing their PB reforms. Problems associated with a top down approach are limited flexibility, bureaucracy and difficulty to gain support from agencies (OECD 2007: 33-34). For an agency, adoption of top-down PB reforms means that it has to deal with the reporting requirements of both its 'regular' agent (e.g. the spending ministry) and the 'administrative' principal (e.g. central budget office).

The Budget process
By linking expected (ex-ante) and measured (ex post) performance to budgeting, PB uses the budget cycle for performance planning and reporting. The logic behind integrating PI and budgeting is to attain an impact of demonstrated results on funding decisions instead of the historical input based approach that dominates government budgeting (e.g. Joyce 2003: 30-

31, Breul&Moravitz 2007: 4-5, Posner&Fantone 2008: 92-93, OECD 2007:11) . As a consequence, performance planning will have to be squeezed into the timeframe of the budget cycle (usually a one year cycle). Because performance planning, especially with regard to policy outcomes often has a longer horizon than 1 year, this has proven to be problematic in many cases. To many in the PB community these problems should not be overstressed as long as performance reporting and evaluation is underdeveloped at many places in government anyway and PB can play a role in further strengthening this capacity and awareness.

In Table 2.1 shows five subsequent phases in the budget process along with their key actors. For the federal US and the Netherlands the actors and the end product of the budget phase are specified.

Table 2.1   Comparison steps in budget process US-NL (partly based on Joyce 2003: 32)

Stage of Budget process (main actors)	Key actors in US	End Product US	Key actors in NL	End product NL
Budget preparation I (Agency, Political Principal)	Agency budget officers, agency subunits	Budget request	Ministry directorates/ subordinate agencies, budget office Ministry,	Budget request
Budget preparation II (Budget Office, Political Principal, Exec. Leadership)	Agency head, agency budget office, OMB, President	President's budget	Budget office Ministry, MoF, Council of Ministers	Draft budget
Budget approval (Political Principal, Exec. Leadership, Legislature)	Agencies, congressional committees, President	Budget resolution; Authorization bills; Appropriation bills (Presid. Veto?)	MoF, Ministers (individually) Parliamentary commissions	Approved budget
Budget Execution (Agency, Budget Office, Political Principal)	Agencies, OMB	Administration of programs	MoF and budget offices Ministries and Ministry	Administration of programs

			directorates/ agencies	
Budget Evaluation (Agency, Budget Office, Political Principal, external auditor/evaluator)	Agencies evaluators and auditors (internal and external), OMB, GAO	Audits and evaluations; recommen-dations	Internal Auditors Ministries, Central Court of Audit, MoF, Parliament	Ministerial Audit reports, Report of Central Court of Audit, recommen-dations

Generally PB systems are designed to affect all steps of the budget cycle and therefore are set to affect all players at least once a year. As PB attempts to integrate the policy cycle into the budget cycle (or vice versa), the moment where funds are allocated is envisioned the main linking pin between the two cycles. This can also be viewed as one of the important limitations of PB as the budget usually has a time horizon of just one year and is informed by past performance while the budget cycle has a longer time-span and focuses on long term outcome projections (Hatry 2008: 296-298)

PB in the Budget Process
When compared to the phases in the section on PB theory (Figure 1.2) the budget preparation phases and the budget approval phase correspond with the *strategic planning* and *budgeting* phases. The Budget execution phase matches the *implementation* phase while during the Audit and Evaluation phase, *assessing performance* takes place (although this does not take place in this phase exclusively). PB theory thus assumes that ideally, after a budget cycle is finished, the lessons from measured performance can be applied to the next round and therefore ought to play a role in the budget preparation and approval phases. Allocative (inter-program) efficiency gains are expected to occur as the budget accommodates shifting funds from ineffective programs to more effective ones. This places a heavy emphasis on the budget evaluation and budget approval phases (Redburn et al 2008: 13, OECD 2007: 66-67).

However, several authors suggest that successes with regard to actual use of PI and therefore the true potential for efficiency gains though PB, lies mainly at the level of the agency during the budget preparation and execution phases of the budget cycle (Joyce 2003: 36-37, Hammerschmid et al 2013: 5, OECD 2014: 4). This corresponds with findings about the actual use of PI by managers for decision making purposes (OECD 2007: 52-53, Wang 2000 as cited in Redburn 2008: 139 , GAO 2004a: 77, 123). Indeed the previous chapter showed disappointing results with regard to allocative efficiency gains at the macro level in the budget approval phase. The assumption that the NPM-ideal type agency model[6]

---

[6] *This NPM-ideal type agency model is considered to be a model based on performance contracting between a principal (often Ministry) and a public agent (often agency). The principal is exercises performance control and*

enhances the performance of public sector organizations is largely untested (Verhoest et al 2012: 4). However there is evidence that increased autonomy and result control can lead to more innovative behavior or an increased use of result oriented management tools within agencies. Both can be considered as preconditions for better performance. (Verhoest et al 2010 in Verhoest at al 2011: 28). Based on these observations this study will primarily focus on operational efficiency gains and result orientation manifested during the budget preparation and execution phases within an agency or in the interaction with its political principal(s). This is illustrated in Table 2.2.

Table 2.2   Focus of research in terms of interaction of budget actors and budgetary phases (based on Joyce 2003: 32)

Budget phase→ Level of interaction↓	Budget preparation (I&II)	Budget approval	Budget execution	Budget evaluation
Intra Agency	**		**	*
Agency-Political Principal	**		**	*
Political Principal-Nat'l Budget Office	*		*	*
Exec. Leadership-Legislature				

** PB expected to have most likely impact on efficiency and result orientation
*PB possibly having an impact on efficiency and result orientation

In Table 2.2 phases in the budget process are contrasted with the hierarchical levels of interaction that occur during this process. The two budget preparation phases from Table 2.1 are integrated into one phase here since the difference between them was in the level of interaction. The four cells that are the primary focus of this study and the five that will have a secondary focus are marked with two or one star respectively.

The evaluation phase can result in valuable information to be used subsequently in the budget preparation and execution phases. Although performance evaluation can be viewed as a condition for obtaining performance informed efficiency gains or result orientation, it provides by no means a guarantee that lessons find their way in measures. As Bob Behn noted, using PI to systematically evaluate effectiveness assumes the presence of a clear coherent mission, strategy and set of objectives combined with a rationalized program

---

used performance evaluation for budgetary purposes while the agent is managed in a performance driven manner.

structure. However in the reality of political life, public sector managers often have to work with conflicting policy objectives, inadequate resources and unreasonable timetables (Behn 2003: 589). In addition many evaluation mechanisms in place in government exist separately from the budget cycle, attending to different frequencies. Nonetheless the results of these evaluation may or may not find their way in budgetary decision making. Therefore the budget evaluation phase will receive a secondary focus in this study.

The interaction between the National Budget Office and the Political Principal may also influence decisions in the principal-agent relationship between Political Principal and agency as was demonstrated by the dialogue on program improvements OMB lead with departments and agencies regarding its PART reviews. However, many intra program reallocations will often be decided on within agencies or in the agreements between an agency and its Political Principal without any interference of the National Budget Office. This is why the interaction between the National Budget Office and the Political Principal will also receive a secondary focus in this study.

The relevance of the interaction between the Legislature and the Government's Executive Leadership for operational efficiency gains and result orientation within an agency is expected to be limited. This is due to the limited role of PI generated by PB systems during budgetary debates with the Legislature as demonstrated with PART (Frisco and Stalebrink 2008: 11) and with performance budgets in the Netherlands (De Vries&Bestebreur 2010: 237) as well as other OECD countries (Curristine 2005: 222). An exception may be the occurrence of an efficiency related incident regarding a Agency in the media with significant political spin-off.

2.2   For what purpose PI is being used

As mentioned in CH 1.1, actual use of PI generated by a PB system is a precondition for intended efficiency improvements to occur. The contribution of PI to inform management decisions within a agency is a vital link to PB theory and has been studied by several authors. Patricia de Lancer Julnes (De Lancer Julnes 2008: 58) points out that proponents of performance measurement traditionally define success in terms of instrumental use (holding someone accountable, make decisions about program budgets or personnel, expanding, cutting back or terminating programs). This often leads to the conclusion that performance measurement has failed. She shows a broader perspective on use of PI by identifying five different uses of knowledge from performance measurement (De Lancer Julnes 2008: 66-70):

- Reassurance: Government shows it is doing what it is supposed to do with taxpayer's money (e.g. reporting PI about cleans streets or safe neighborhoods)
- Compliance: Agencies demonstrate that they comply with performance measurement regulations (e.g. schools report test results to receive funding)

- Program learning: Learning from results may lead to program changes or maybe just to a better informed dialogue.
- Enlightenment: Externally, enlightenment can lead to mobilization and put an issue on the political agenda (e.g. awareness of waiting lines in public service). Internally, enlightenment can lead to more informed decisions and better educated stakeholders. This can generate new insights and challenge previously held perceptions.
- Legitimization: PI can be used to rationalize, justify or validate current, past and future course of actions and decisions (e.g. justifying budget requests, or defending oneself towards critics)

Looking at these uses of PI and the classification of potential contributions of PB mentioned earlier, the use of PI for the purposes of program learning and enlightenment seem to be relevant for both operational efficiency and internal result orientation. Reassurance seems to address the transparency aim of PB. Success of PB regarding transparency will not be the focus of this study as explained in CH 2.1.

Using PI to legitimize existing positions like budget claims is a phenomenon easily recognized by practitioners. As this merely adds to existing dynamics, use of PI for legitimizing purposes is doubtful as a claim to success of PB when looking at the contributions PB reformers intended to make. It is however possible that performance data, intended to legitimize a position, helps generate conflicting data from others and thus becomes a dominant factor in an advocacy based dialogue with other stakeholders in the policy process. In this case legitimization could contribute to operational efficiency and internal result orientation through enlightenment. Nevertheless it does not seem likely that the existing dynamics will shift easily as a result of adding PI alone, especially when this information is framed to serve the interest of the party reporting it. Compliance with PB regulations, whether this happens because of financial incentives or safeguarding a reputation (naming and shaming) can, for the purpose of this study, not be counted as PB success either. Doing so would indeed regard PB reform as a self-legitimizing exercise.

So the role of performance information, when expected to contribute to efficiency and result orientation, would foremost be to contribute to program learning or enlightenment. Drawing a parallel with common literature on learning organizations, the difference between using PI for program learning and for enlightenment resembles the difference made between single loop and double loop learning respectively. According to the classic work of Chris Argyris with regard to learning organizations, single loop learning involves the detection and correction of errors given a particular goal or plan. Double loop learning involves questioning the underlying reasons and motives behind the facts such as norms, values and objectives (Argyris 1994: 78-79). When studying the lessons the NL government learned from policy evaluation, Peter van der Knaap applied a similar difference between

corrective learning and fundamental learning, the latter adding to the existing underlying policy assumptions (Van der Knaap, 1997:57-58).

The use of PI for these purposes may eventually result in decisions that fit the more traditional definitions of PB success like *holding someone accountable, make decisions about program budgets or personnel, expanding, cutting back or terminating programs*. This may occur at different levels of interaction. Following a slightly different classification of PI use, Moynihan arrives at a similar conclusion when talking about *purposeful use of PI*, meaning that public employees use data to improve program performance through goal-based learning that gives rise to efficiency improvements, better targeting of resources and more informed strategic decisions or by tying indicators to rewards/sanctions in contract arrangements (Moynihan 2009: 588). Subsequently these performance informed decisions are supposed to contribute to more efficiency. This is the next link in PB's micro model.

2.3   How PI is being used by an agency

Disregarding the use of PI for other purposes than targeting efficiency and effectiveness improvement is an important distinction as it narrows down the search for successful PB implementation to the use of PI by ministries and agencies to actually increase effectiveness and efficiency. When exploring this question in even more detail, purposeful use of PI can be associated with a set of specific types of PI use that are likely to eventually result in budgetary consequences.

 In the previous sections it was stated that the use of PI for purposes of organizational learning and enlightenment can lead to a better educated dialogue and performance informed decisions. At the agency level, this is expected to be able to contribute to greater operational efficiency and result orientation, most notably in the budget preparation and execution phases. Just to get a more accurate picture of how PB, according to its proponents, exactly contributes to increased efficiency, this paragraph further analyses the role that PI is expected to fulfil in the process of budgetary and performance planning and monitoring within a agency and in the interaction with its principal(s).

The potential for the use of PI in the different phases of the U.S. federal budget process was explored by Phil Joyce for the budget preparation phases, the agency can use PI to (Joyce 2003: 37-42):
-    To justify budget requests
-    To strategically re-allocate internal resources
-    To determine productivity of different agency components
-    To determine overlapping services
-    To decide on outsourcing decisions (make or buy)

During the budget execution phase the agency can use PI to (Joyce 2003: 51-59):

- Understand the impact of external events on agency performance goals
- Allocate funds internally
- Allocate funds to third parties
- Monitor cost and performance
- Motivate staff to act consistent with agency goals

The evaluation and audit phase, when including performance evaluation, can be expected to mainly contribute to decision making in an indirect way. This is because favorable conditions for direct are usually absent in this phase. These favorable conditions (that are usually absent) can be summed up as: clear and unambiguous results that are relevant to an issue up for decision and are understood by decision makers who are interested (Weiss& Bucuvalas 1980 as quoted in Van Nispen 1993:37).

When asking government personnel what they use PI for, those who claim to do so mention a wide variety of things. When studying of the use of PI by U.S. counties, Wang found that the majority of counties that used performance measurement, used the results for (Wang 2000 as quoted in Redburn 2008: 139):
- Preparing departmental requests (78%);
- Identifying service problems and evaluation (63%) and
- Analyzing funding levels (58%)

According to a 2003 GAO survey of U.S. federal managers the following management activities were mentioned as primarily involving the use of PI (GAO 2004a: 77)[7]:

a)   Allocating resources (60%)
b)   Setting individual job expectations for staff (60%)
c)   Rewarding staff (60%)
d)   Setting program priorities (59%)
e)   Setting new, or revising performance measures (58%)
f)   Adopting new program approaches or changing work processes (56%)
g)   Refining program performance measures (51%)
h)   Coordinating program efforts with other internal or external organizations (49%)
i)   Developing and managing contracts (41%)

Comparable results were found at the U.S. state level by Melkers & Willoughby with the activities of reporting and strategic planning as notable additions to the uses mentioned so far (Melkers & Willoughby 2004 in Breul & Moravitz: 88).

---

[7] The % mentioned present the 2003 percentages of those managers who indicated to use PI to a great or very great extent

A few conclusions can be drawn when looking at the ways in which PI is supposed to be used by an agency according to PB theory, it's use in reality and relating this to the purposes for which the information is used. Firstly, the potential use of PI in ways that, theoretically, can contribute to efficiency and result orientation are broadly similar to the actual use. At least according to responses obtained from surveying agency staff. This suggests that in cases where PI is used by an agency, PB theory might indeed work as envisioned. This does not tell anything about whether the intended results, namely efficiency gains and increased result orientation, actually do occur. Besides, the fact remains that when PI does not get used, PB still cannot be expected to work. It is unclear how widespread the use or PI in these studies is and if so, how intensely or frequently. Practitioners can usually, without too much effort, name plenty of cases where this has not been the case.

Secondly, to result in operational efficiency gains and increased internal result orientation, PI is used for the purposes of *program learning* and *enlightenment.* A number of activities that involve PI use seem to contribute in an indirect way at best to the end-goals of operational efficiency and result orientation. Justifying budget requests in the budget preparation phase refers to *legitimization* purpose of PI use. Preparing requests does not add to the purposes of PB either, except when it includes efficiency gains resulting from the other activities mentioned. Setting, revising or refining performance measures can be about *program learning* but in a rather instrumental way. When it is more about improving the PB system itself than about its intended results, namely program efficiency, it may actually refer more to the *compliance* purpose of PI use. As noted previously, having better or more performance measures can hardly be a goal at itself. The use of PI for reporting activities to increase accountability purposes can be associated with *reassurance* or *compliance* purposes more than with direct *program learning*. However reporting PI (or having to report it) may indirectly contribute to *enlightenment* either within an agency (internal accountability) or up the agency's accountability chain (external accountability). Finally, using PI for the activity of strategic planning, just like preparing a budget request, may or may not contain lessons that improve efficiency from the other activities mentioned.

In summary, four categories of activities stand out in which the use of PI can be considered as purposeful PI use by an agency and is likely to contribute to operational efficiency or internal result orientation:

1. Major policy shifts and paradigm changes in public opinion can occur because of the insights PI can offer to vital stakeholders. Internal or external policy specialists can use PI to highlight successes or failure. If they use this information to convince other stakeholders, this can be a **driver of major policy shifts**. Replacing or adjusting unsuccessful policy measures is expected to eventually lead to better overall program performance and efficiency.
The intention of the use of PI in this case is *enlightenment.*

Agency activities using PI that can contribute to this are:
- Performance reporting for external accountability
- Setting program priorities
- Strategically reallocate internal resources
- Understand the impact of external events on agency performance goals

2. PI can be a **major tool in managing principal-agent relationships**. Performance reporting is crucial as a solution to the principal's information asymmetry problems as he can get an accurate picture of the relative performance of an agency and the costs associated with its service delivery. This way a principal can negotiate better deals with its agent(s) leading to more efficient service delivery. The intention of the use of PI in this case is *enlightenment* as well as *program learning.* Agency activities using PI that can contribute to this are:
- Deciding on outsourcing decisions
- Developing and managing contracts
- Monitor cost and performance and contract management
- Allocate funds to third parties

3. Availability of PI **increases an organization's learning capacity** as it provides the results of previous efforts. Evaluative analysis of results and internal dialogue about relevant lessons through learning forums can be expected to lead to better decisions and thus promote efficiency. The intention of the use of PI in this case is *program learning*.
- Coordinating program efforts with other internal or external organizations
- Analyzing productivity and funding levels
- Allocating internal funds
- Identifying service problems and changing work processes
- Adopting new program approaches following evaluation

4. PI can be used to **motivate staff** to act in accordance with the goals of the agency targets and the principal's strategic goals. This can be done by adopting formal performance management incentives but even without these being in place, repeated communication of vital results by agency leadership may have an impact. Increased result orientation of an agency is intended to indirectly support *enlightenment* and *program learning* as it is meant to unlock the organization's learning potential. Agency activities using PI that can contribute to this are:
- Motivate staff to act consistent with agency goals
- Setting individual job expectations for staff
- Rewarding staff

The expected contribution of PB to efficiency and result orientation through purposeful PI use is summarized in tables 2.3 and 2.4. Combined they form a micro model of PB.

Table 2.3   Micro model of PB: Where to look for success

PB is expected to lead to more	Manifested during	In the interaction levels
- Operational efficiency - Internal result orientation	- The Budget preparation phase - Budget execution phase + possibly: - Budget Evaluation phase (indirectly)	- Intra agency - Agency vs. Political Principal + possibly: - Political Principal vs. National Budget Office

Purposeful use of PI can be associated with a set of specific types of PI use that are likely to eventually result in budgetary consequences as illustrated in Table 2.4.

Table 2.4   Micro model of PB: purposeful PI use with potential impact on efficiency and effectiveness (based on GAO 2004b, Joyce 2003, Melkers & Willoughby 2004, Moynihan 2008).

Specific type of PI use for program learning and enlightenment…	…may result in effectiveness and efficiency gains through:
- Performance reporting for external accountability - Setting program priorities - Strategically reallocate internal resources - Understand the impact of external events on performance goals	Goal alignment or adapting policy assumptions
- Deciding on outsourcing decisions - Developing and managing contracts - Monitor cost and performance and contract management - Allocate funds to third parties	Better resource management
- Coordinating program efforts with other internal or external organizations - Analyzing productivity and funding levels - Allocating internal funds - Identifying service problems and changing work processes - Adopting new program approaches following evaluation	Taking corrective actions

- Motivate staff to act consistent with government agency goals	Goal orientation and motivating staff
- Setting individual job expectations for staff	
- Rewarding staff	

Successful PB implementation at the agency level can be expected to show similarities with the model presented above. Now that a micro model of PB implementation has been defined, we know what successful PB would look like and where to look for success. Whether these dynamics actually occur is dependent upon a lot of factors that most PB theorists seem to have left out of their analysis. In the final part of this chapter, institutional and behavioral insights will be brought into the picture as a possible explanation for the use of PI to promote efficiency. First however we return to the research question introduced in Chapter 1.4 that can be further refinement using the model introduced here.

2.4   Refining the research question

In the previous sections PB theory with regard to its intended result of increasing operational efficiency and internal result orientation was refined to include more precise expectations of what successful PB would look like and where to look for it. We now know in which phases of the budget cycle and at what levels of interaction between its actors performance budgeting is most likely to deliver results. We also have a clearer understanding in which ways performance information, generated in a PB system, can be expected to contribute to more operational efficiency and more internal result orientation. Having gained more of a micro-perspective on PB theory, Schick's observation that *'governments that don't manage for results will not budget for results'* (see CH 1.4) deserves a second look (Schick 2003: 102).

If it is true what Schick says, introduction of PB by an agency or its principal(s) would only deliver results if a number of other, independent factors, are present. As a consequence, PB successes that are reported from an agency or their principal(s) would, according to Schick, only have been made possible because *'managerial arrangements that make results paramount'* were already in place or because PB was accompanied by other *'transformations in public management that enhanced performance'* (Schick 2003:102). The fact that internal result-orientation of an agency has itself, also been reported as a consequence of adopting a PB system, points at a possible fallacy of circular justification regarding successful PB adoption: Does successful PB implementation **create** favorable circumstances for purposeful PI use? Or can successful PB implementation **only result from** favorable circumstances for purposeful PI use? We therefore ended Chapter 1 with the general research question:

Are result orientation of a government agency and operational efficiency gains achieved through PB?

Finding an answer to this question can only proceed after the favorable circumstances for result orientation are identified and defined. The question what *managerial arrangements* and *transformations* are regarded necessary for PB to work remains undisclosed by Schick. He does refer however to the politicians and managers as actors in the budgetary process relevant factors to the results of PB. The focus of Schick on arrangements that result from transformations and reforms is somewhat surprising. An underlying fundamental question is whether centrally guided reform initiatives such as PB did make a lasting contribution to government agencies being transparent about their goals, measuring their results and using these results to learn and improve. While some early evaluations suggest that this was the case, recent studies increasingly assert this view and hint at a more complex reality (Dull 2009b: 2-3, Moynihan& Lavertu 2012: 598, Posner& Mahler 2012: 3) as this quote illustrates:

> '*Top-down reforms always struggle, but they have the best opportunities for success when they amplify or encourage preexisting patterns of behavior. The same is likely to be true of the success of performance management routines. In organizations in which there is already some interest in performance management, reforms to encourage use will be more likely to succeed because they will face little opposition or link to preexisting operational routines (Moynihan& Lavertu 2012: 601)*

The importance of a fit between new routines of PI use called for by PB reforms and existing organizational routines, should make an institutional perspective an inevitable part of any exploration of purposeful PI use. Building on the understanding of the micro model of PB and where to look for PB successes, the central research question can be further refined to this central research question:

How do underlying cultural and historical factors explain successful PB in government agencies?

To answer this question several sub questions have to be answered regarding a case of successful PB implementation in the relationship between an agency and its political principal:

<u>Sub question 1</u>: *What is considered a case of successful PB in a government agency?*

Since we are seeking to identify purposeful use of PI from a PB system in an a principal-agency setting we can identify between use of PI in the interaction between the political principal and the agency and internal use of PI within an agency. Summarizing the specifications made so far, a 'textbook example' of PB implementation in line with the assumptions of NPM and agency theory is assumed combine four characteristics:

- *PI is used in the budgetary process between agency and principal in addition to traditional budgeting*
- *An intense performance dialogue exists between agency and principal*
- *The agency has an ambitious and sophisticated PB system in place*
- *PI from this system used internally by the agency for budgeting and performance management*

If these four characteristics are found to be present in a public sector agency and the interaction with its principal, PI is indeed used for budgetary purposes and may be expected to positively influence effectiveness and efficiency alongside micro model of PB introduced in Section 2.3. Before being able to test and compare these characteristics in real life cases, they will be further developed into more specific into indicators in Chapter 4.

Sub question 2: *What are considered relevant underlying historical and cultural factors?* Answering this question will require a further analysis into institutional characteristics that are likely to be autonomous from PB implementation. Chapter 3 and 4 will identify relevant factors that may offer an explanation for purposeful PI use.

Sub question 3: *Does the presence of these cultural and historical factors offer a convincing alternative explanation for successful PB?*

This can only be answered if the results from sub questions 2 and 3 are tested empirically. In case of a positive answer there should be serious doubts about the added value of introducing a PB system because in those cases where PB did work, PI was likely to be used for planning and budgeting anyway. If not, introducing a PB system may have contributed to integrate PI into planning and budgeting and that way contributed to efficiency and result orientation. The approach to address these cases and the answer the general questions will take some further refinement and will be disclosed in Chapter 4. After the results of the empirical part of this research are presented (Chapter 5-8), an answer can be provided to the central research question.

As for now, the relevant underlying cultural and historical factors remain unidentified. To answer this refined research question, the conditions that influence the use of performance information, have to exist separately from adoption of the PB system itself. The quest for assessing the results of PB so far has revealed that the genuine success of reforms in public budgeting may quite heavily rely on 'soft' behavioral factors like leadership, culture and other autonomous factors that may be entirely out of the control of budget departments, their staff and their political leaders. The influence of these factors on the use of PI and therefore the success of PB will be addressed in this chapter's last paragraph.

2.5   Explaining the use of PI differently

The actual use of PI by an agency and its principal(s) for the purpose of enlightenment and program learning proved vital for PB to work according to theory as was demonstrated earlier in this chapter. It is by no means self-evident that performance information, once it is generated and disseminated, will be used in accordance with this refined PB assumption. The problems encountered with the use of PI by public sector organizations are summed up well by Moynihan in his Interactive Dialogue Model of PI Use (Moynihan 2008: 102). This model challenges PB theory at the micro level. The model's central assumptions are:

1. PI is not comprehensive,
2. PI is ambiguous
3. PI is subjective
4. Production of PI does not guarantee use
5. Institutional affiliation and individual beliefs will affect selection, perception, and presentation of PI
6. The context of dialogue will affect the ability to use PI to develop solutions

The first three assumptions reflect the idea that PI usually consists of one or more indicators of actual performance and as such never reflect a whole universal truth about performance, the existence of which can obviously be debated in the first place. For example, successful liberalization of a taxi industry in the Netherlands can be reflected by a number of conflicting indicators like prices for cab-rides, profit margins for taxi-entrepreneurs, pay for taxi-drivers, customer satisfaction, availability of taxis etcetera, depending on one's perspective. The first three assumptions partly mirror the 'misassumptions' of PB as by Radin (Radin 2006 see CH 1.2). The importance of these notions lies in the fact that once you accept that PI is not comprehensive, ambiguous and subjective, the rational basis for a performance informed dialogue between actors in the budget process, vital to PB theory, is largely gone. What remains is a dialogue in which behavioral and institutional factors play a large role and are even likely to dominate.

The fourth assumption deals with problems regarding the use of performance data. Moynihan also notes that when PI is used, it is not always used in the way PB reforms envisioned it to be used (Moynihan 2008: 95) Since we focus on how PB theoretically is supposed to lead to efficiency gains, we disregarded the use for purposes of reassurance, compliance and legitimization. So, not only does production of PI not guarantee use, use itself may or may not contribute to efficiency depending on the purpose of use. Moynihan's first four points are largely covered by the previous sections with regard to the potential contribution of PI to efficiency.

Taking the fifth assumption a bit further, the consequences of institutional affiliation and individual beliefs may well exceed the selection, perception and presentation of PI It is not

40

hard to imagine that these factors will offer explanations of whether the efficiency gains according to PB theory do occur or not. It is true that PI is the central driver of the process that, according to PB theory, will result in efficiency gains. However it is possible to assume that selection, perception and presentation of PI, even if this is geared towards enlightenment or program learning, still does not result in the efficiency gains that PB theory expects. This is so because the conditions for internal learning may reflect only a small part of actual organizational behavior. There may be other organizational factors that may form impediments to actually harvesting on the performance informed lessons.

The context of dialogue presented in the sixth assumption may itself be heavily influenced by institutional affiliation and individual beliefs as well. The way the organizational dialogue is conducted between those who measure performance and those who account for the performance measured can prove an obstacle to applying the lessons to be learned. A dialogue that involves a few of parties that maintain a close relationship offers the best guarantees for efficient exchange of information because much information can remain implicit because they are more homogeneous in several aspects (Moynihan 2008:19). PB can be viewed as an attempt to break open the policy dialogue to be able to involve others (the financial department, parliament) by making many implicit assumptions explicit. If a perceived outsider is forcing his insights into a policy dialogue, defensive routines are likely to occur on the part of the agency. This can result in a ritual dialogue in which referral is made primarily to a formal positions and documents rather than actual policy content (Van der Knaap 1997: 260-262).

Attention to the influence of institutions on political and social outcomes is known as neo institutionalism and can be viewed as a wider trend of applying social theory to economic rationalist assumptions. What Moynihan expresses in his model is in fact an institutionalist critique on performance budgeting, whose very roots can itself be traced back to new institutionalism. Because the institutional context is a likely factor that could explain why PB does or does not result in the intended effects highlighted here, the next chapter will turn to neo-institutionalism and some of its schools of thought. In the next section neo institutionalism and the principal agency problem will be explored as they, in particularly bear relevance to the theoretical core for this research.

CHAPTER 3   AN INSTITUTIONALIST PERSPECTIVE

In 2010 a Chinese newspaper reported about the practice of publicly humiliating criminal offenders in the provinces of Hunan and Guangdong. This humiliation consists of parading criminals through the streets in prison outfits with signs around their neck explaining their crimes (be it theft, murder or prostitution). Large amounts of people, including the convict's children, are forced to watch these parades and TV stations and newspapers are obliged to report on them. Although outlawed by Beijing, local politicians, media and law enforcement officials continue to support public humiliation campaigns. One party spokesman from the city of Loudi defied critics by explaining: _this method may not be humane but it works well as deterrence and demonstrates results. Therefore this method is correct_ (Source: Article p.7 NRC next Aug 18th 2010). The debate about whether behaviour is shaped by following rules about what is considered appropriate or by simply rewarding what works well is an institutional one that divides many public policy issues. In this chapter these different approaches are explored and applied to the phenomenon of PB and its ability to shape and control agency behaviour.

3.1   Introduction

In Chapter 1 it was argued that attempts of governments worldwide to structurally integrate their budgets with performance planning and reporting has, so far, not brought all of the improvements that were envisioned by many reformers. As explained in Chapter 2, a multitude of promises and expectations was communicated at the start of Performance budgeting (PB) reforms. Of the different criteria for success two stand out as having rather inconclusive evidence and are therefore in particular deserve further exploration. These are the claims of PB contributing to efficiency gains from operational allocation and increased internal result orientation. Evidence suggests this is most likely to appear during the budget preparation and budget execution phases. In order for PB theory to work during these budget phases as envisioned, PI is expected to be used in the principal-agent relationship for purposes of enlightenment and program learning.

The question whether the dynamics to support the working of PB this way are present, is something that most PB theorists seem to have left out of their analysis. According to Moynihan's Interactive Model of PI Use, these dynamics are largely shaped by institutional affiliation and individual beliefs as well as the context of the dialogue (Moynihan 2008: 102-113). In line with Moynihan's neo-institutional critique on PB theory, some relevant theoretical background will be explored to enrich the understanding of the dynamics underlying organizational behavior. These institutional insights will later be applied to assumptions regarding PI use.

The linking of theoretical concepts treated in this chapter is schematically displayed in Figure 3.1. Reading the chapter is probably necessary to make this figure make more sense. It does

reveal however that PB's assumptions towards PI use are associated with one neo-institutionalist school of thought whereas the other two schools might offer alternative explanations for (a lack of) PI use.

Figure 3.1   Linking of Theoretical concepts
              * IMPIU: Interactive Model of PI Use
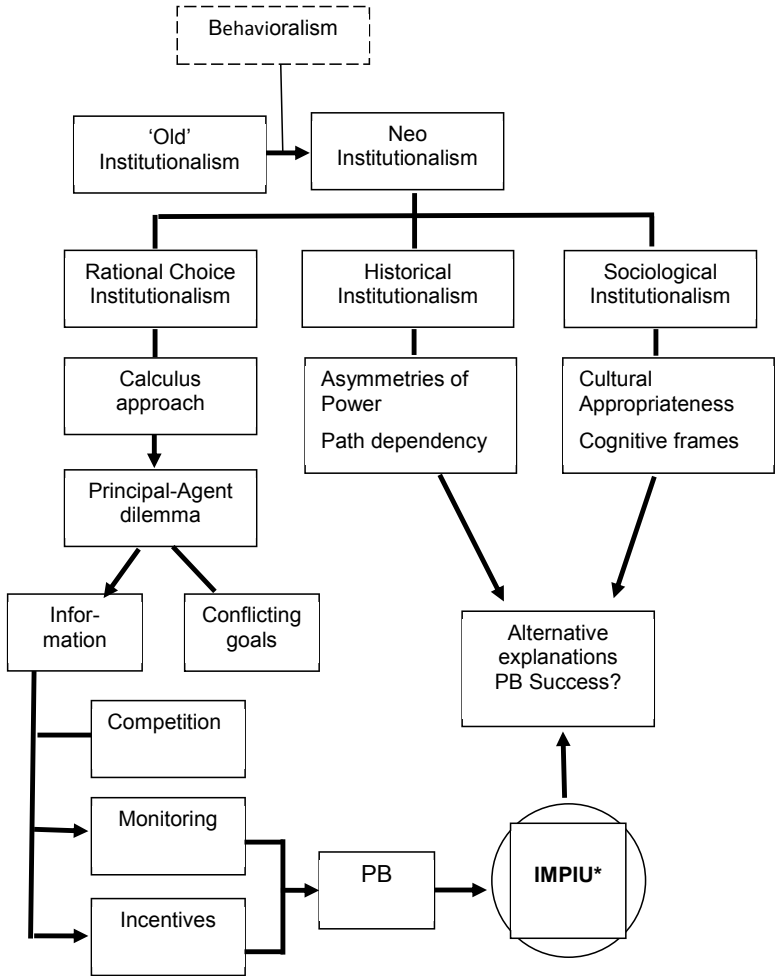
3.2   The case of 'old' institutionalism: organizations matter

Institutionalism can be traced back to the 18th and 19th century as some of its key notions were already discussed by classical philosophers in those days. For example, Rousseau stressed the difference between the public interest and the aggregation of individual interests and De Tocqueville saw in local political institutions an antidote for despotism (Immergut, 1998: 8-17).

An institutional approach in British or American studies of the 20th century usually refers to investigating the complex impact of political systems on certain outcomes not commonly associated with these outcomes. The emphasis on these studies was more on description and normative evaluation and less on explanation or theory building. Insofar 'old institutionalism' turns to explaining phenomena; this tends to be done by focusing on formal rules of institutions (Bell 2002: 2, 13). An institution in these studies is an institutional arrangement such as for example a parliamentary or presidential system of government. Often institutions can also refer to the actual organizations that result from these arrangements. Old institutionalism is not dead as describing the impact of formal institutional arrangements can still shed some welcome new light on long standing problems. With regard to budgeting, contemporary examples of studying the impact of institutions in the classical sense are: studying the impact of legislatures on budgeting by Paul Posner and Chung-Keun Park (Posner&Park 2008: 2-5), the relationship between country's governing situation and budget deficit reduction (Allen Schick 1993: 187-235, edited by Weaver & Rockman) or possible bias of a Republican Presidency on the outcome of performance measurement (Gilmour & Lewis 2006: 746).

In their 1993 work *Do Institutions Matter? – Government Capabilities in the United States and Abroad*, Weaver and Rockman identify three tiers that explain differences in government capabilities:

First Tier institutions: Presidential or parliamentary system. This affects the decision making system with regard to the level of party discipline, the recruitment of executives and the degree of centralization of power and accountability.

Second Tier Institutions: Regime type or Government type. This affects the decision making process with regard to a number of decision making moments (veto points), the stability and cohesion of the government elites, their autonomy form electoral and constituency pressures and the level of influence of interest groups.

Third Tier Institutions: To this 'leftover category' belong broad framework political institutions such as judicial review, federalism and bureaucratic autonomy and secondary institutional characteristics like voting rules and bicameralism. Besides these 'classical' institutional factors, Weaver and Rockman also include a number of other factors in their third tier category. Political conditions, policymaker's goals, socioeconomic and demographic conditions and past policy choices reflect 'an even broader array of non-institutional influences on capabilities' (Weaver & Rockman, 1993: 7-37). After evaluating 10 cases, they conclude that although institutions do affect government capabilities, their effects are contingent (Weaver and Rockman 1993: 446).

3.3   Perspectives in the institutional debate

Although Weaver and Rockman recognize there is a 'broader array of influences' relevant to government capabilities, they deem these as non-institutional. This is a view not shared by

all institutionalist authors. Expanding the definition of an institution beyond the formal arrangement or organization, an institution can be viewed as a relatively stable collection of practices and rules defining appropriate behavior for specific groups of actors in specific situations. That way, institutions socialize individuals by legitimizing behavior associated with roles to be enacted or sanctioning behavior that differs from what is considered appropriate. Practices and rules are also embedded in resources and the principles of their allocation (March&Olsen 1998: 984). A more simple definition of institutions was given by William Ryker: rules about behavior, especially about making decisions (Ryker as cited in Ostrom 1986:1). Institutionalization refers to the emergence of institutions and individual behaviors with them. An institutional approach is one that emphasizes the role of institutions and institutionalization in the understanding of human actions within an organization, social order or society (March&Olsen 1998: 984).

At the heart of the institutional debate is a central notion on the drivers of human behavior: Are actions driven by expectations of consequences, also referred to as the 'logic of consequence' or are actions driven by rules that come with a particular role or identity, also referred to as 'logic of appropriateness'? The answer to this question may affect one's view on how political life is organized, to a considerable extent (March&Olsen 1998: 952).

From a *'logic of consequence'* viewpoint, organization of political life is a result of actions by rational self-maximizing actors with conflicting interests and varying resources that negotiate a set of contracts. These actions are consequence based and are of an instrumental nature. Although theories of bounded rationality have modified this perspective somewhat, the central notion that behavior is primarily driven by expected consequences that will maximize one's benefits is left untouched (March&Olsen 1998: 950).

From a *'logic of appropriateness'* point of view, political life is a result of rules that are socially constructed, publicly known, anticipated and accepted. These actions are rule based and reflect a perceived identity of the actor. The central notion here is that people are role players and rule followers that will act out of duty or social pressure (March&Olsen 1998: 952). Behavior is therefore primarily driven by what people feel is appropriate behavior.

These two logics are not mutually exclusive. Most actions are motivated by elements of both logics as decision makers tend to follow rules and calculate consequences at the same time (March & Olsen 1998: 949-954). A decision that makes sense from both logics may be more likely to be taken and implemented. This would imply that strategies based on maximizing benefits will be more successful when these strategies are considered appropriate by those involved. Vice versa, strategies that fit institutional rules and expectations may be more successful when individual benefits can be expected as well.

About the interaction between the two logics with other principles, several classifications have been suggested. March & Olsen refer to the issue of historical efficiency as another debate that divides institutionalism. At the one hand there are instrumentalists who view history as efficient because it rewards those who successfully compete for survival. Players are set to reach predetermined optimal equilibriums regardless of their specific characteristics other than preferences and resources (March & Olsen 1998: 954, March & Olsen 1984: 737). On the other hand there are those who view history as inefficient. They argue that pressure of survival is neither constant nor precise enough to influence outcomes and point out that an equilibrium may not exist or will never be reached. The latter can be the result of history responding more slowly than changes in the environment leading to new equilibria. Instead, the inefficient history perspectives stress an institution's origin, history, and internal dynamics as influencing eventual outcomes. (March & Olsen 1984: 737, March & Olsen 1998: 955).

Combining these two institutionalist divides results in four perspectives on the dynamics of political order as shown in table 1. Studies in the *functional rationality* perspective assume calculating individuals striving for individual advantage. Preferences and interests are hardly affected by politics or institutions. Studies in the *history dependent rationality* perspective show that consequential history is unpredictable and path-dependent. History coevolves with its consequences.

Following the logic of appropriateness, *functional institutionalism* stresses that rules, norms and identities shape behavior. Those rules however, are a product of an efficient history in which only those institutions survive, that are most fit to their environment. Just like functional institutionalism, *history dependent institutionalism* assumes that rules, norms and identities co-evolve with the worlds in which they act, but do so in an unpredictable, though path dependent way (March & Olsen 1998: 956-958).

Table 3.1   The logic of consequence and the logic of appropriateness coupled to two the notion of historical efficiency (Source: March&Olsen 1998: 957)

	Efficient History	*Inefficient History*
Logic of consequence	Functional rationality	History- dependent rationality
Logic of appropriateness	Functional institutionalism	History- dependent institutionalism

When trying to place NPM and PB's body of thought in this framework, PB with its underlying assumptions seem firmly embedded in the logic of consequence, PB shares elements of both functional rationality as well as history dependent rationality. As much as

PB theory assumes that individuals are motivated by individual advantage, it does not assume that the outcome of maximizing civil servants' individual benefits has resulted in survival of the most efficient institutions. To the contrary PB and especially NPM are set to offer tools to steer the inefficient government inherited from history, towards more efficiency by using incentives that appeal to individual advantage. The PB analysis shares the conviction of history dependent rationality that the absence of a predictable pressure for survival has resulted in inefficient institutions. Installing competition among public service providers is therefore commonly presented as another approach to solve this problem in addition to PB.

3.4   Neo institutionalism

Neo institutionalism is neither a theory nor a consistent critique of one. It just argues that the organization of political life makes a difference (March & Olsen 1984: 747). As mentioned earlier, the roots of neo-institutionalist concepts can be traced as far back as the 18th century. It can therefore be argued whether the term neo institutionalism is justified (Immergut, 1998: 8-17). Others stress that while new and old institutionalism are not identical, neo institutionalism blends elements of old institutional into non-institutionalist styles of recent theories of politics (March&Olsen 1984: 738). Where 'old' institutionalism primarily analyzed political institutions and their formal relationships, after World War II, the conviction grew that true political power is obtained by informal relationships within and beyond government institutions. The behavioral approach tried to fill this gap by explaining 'phenomena of government' by observable behavior. Observed behavior was considered to indicate the true preferences of persons. Neo institutionalism[8] is commonly viewed as a response to the political behavior movement in the 1950s and 1960s. The critique of neo institutionalism on behavioralism, centers around three main points (Immergut 1998:6-8)

- Behavior does not reveal preferences
- Preferences cannot be added up to become political preferences
- Institutions are biased and collective decisions are normative

Behavior does not reveal preferences[9]
A vital notion that is implicit in both behavioralism and institutionalism is that an individual's real preferences are hidden. To behavioralism, real preferences can be revealed by observing behavior (such as voting). The behavior displayed gives away the real preferences. Institutionalists add to this notion that individuals may opt for displaying other preferences than their real ones. This can be the case because people choose to satisfice rather than

---

[8] Both the terms *institutionalism* and *neo-institutionalism* may at times be used in this section to illustrate institutionalist viewpoints used in the theoretical model. This is regarded tolerable as the borderline between both terms can be debated because of the continuity of elements shared by both. In addition the debate on the proper terminology is not felt to add sufficiently in illuminating the conflicting assumptions on agency behavior with regard to PI use.
[9] The difference between stated and revealed preferences is more widely acknowledged as a source of inaccurate research results in amongst others economics, marketing and sociology.

optimize for reasons of feasibility or because conflicting preferences force people to prioritize between them. There may be a difference between displayed and real preferences for exactly these reasons. Thus neo institutionalists argue that political behavior is insufficient for explaining phenomena of government for behavior can only be understood in the context of institutions in which it occurs. They claim that behavioralists fail to see that different circumstances may result in different displays of behavior by the same person, whose interests may not be clear in the first place. This is why exposure and analysis of discrepancies between potential interests and those expressed in political behavior are important to institutionalist theory.

Preferences cannot be added up to become political preferences
The behavioralist notion that individual behavior preferences or interests can be aggregated into political preferences is rejected by neo-institutionalists as far too problematic. Reasons for this are the complex nature of human interests themselves and the fact that the mechanisms for aggregations (e.g. discussion and negotiation) reshape the interests and preferences themselves. Because the summation of preferences is considered problematic, aggregated preferences cannot fully explain collective decisions. This notion is supported by theories of group behavior by public choice authors like Mancur Olson who demonstrated that large groups usually have no incentives for furthering collective interests (Olson 1965: 48).

Institutions are biased and collective decisions are normative
Because political behavior and collective decisions are normative by nature, institutional configurations may privilege particular sets of interests and may need to be reformed. Recognizing a bias in institutions and their decision making processes suggests that political behavior is shaping decision making processes and their outcomes.

The institutional critique on behavioralism stated that displayed behavior poorly reflects real political preferences and includes institutional biases that may or may not be known. This means observed behavior offers a poor explanation for the phenomena of government. Central to neo institutionalism are a number of phenomena that are easily observed but often hard to explain (March & Olsen 1984: 747).

3.5   Different branches of neo-insitutionalism

At least three branches of neo-institutionalism have been generally recognized. They all developed their own reaction to behavioralism by blending institutional approaches with existing political theories. These three schools are: Historical Institutionalism, Sociological (or Organizational Theory) institutionalism and Rational Choice institutionalism. The first two are shared jointly under Historical Institutionalism by some (Bell 2002: 5-8). Although Historical and Sociological Institutionalism share the logic of appropriateness, distinct differences can be made between both.

Out of criticism on the ability of these three branches of institutionalism to explain institutional change, discursive or constructivist institutionalism emerged in recent years as a fourth branch of institutionalism (Schmidt 2008: 304-305, Hay 2006: 61). Discursive institutionalism makes the point that political discourse is undervalued as an explanation for policy change or continuity. Contested by some for not moving substantially beyond historical institutionalism (Bell, 2011:1), this variant, as well as other forms of neo-institutionalism that are distinguish by some authors, will be left out of the analysis in this chapter. The reason for this is that the seemingly continuous debate on the proper classification between neo institutional schools of thought goes beyond the deepening of the understanding of the interpretation of PI use sought after in this chapter. Therefore, the origin, characteristics and central concepts of only the three earlier mentioned variants will be shortly addressed.

Rational choice institutionalism
This branch of neo institutionalism originally sought to explain stability of outcomes of Congressional legislation given the instability of preferences and majorities. This was done by stating that institutions lower transaction costs for Congress and solve many of the collective action problems. Emphasis is placed on transaction costs, property rights and rent-seeking. Although rationality is bounded, individual goals are achieved through institutions. Institutions are systems of rules and inducements. An especially relevant concept to rational choice institutionalism is agency theory. Organizational structure is explained by the way it minimizes transaction, production or influence costs. Rational choice explains the relationship between institutions and individual behavior with a highly generalizable set of concepts but with a simplistic image of human motivation. Politics are explained by management of uncertainty. Rational choice institutionalism demonstrates the importance of information flows for power relations and political outcomes as well as strategic interaction (Hall & Taylor 1996: 10-14).

Historical Institutionalism
Historical institutionalism is a response to group theories of politics and structural functionalism and developed notions of how, in the competition for scarce resources, some interests were privileged and others demobilized. Many historical institutionalists no longer saw the state as a neutral broker among competing interests but as a complex of institutions capable of structuring outcome of group conflict (Evans 1985 quoted in Hall & Taylor 1996: 6). Historical institutionalism stresses asymmetries of power, path dependency, unintended consequences and inefficiencies of existing institutions.

Historical institutionalists divide history in periods of continuity altered by 'critical junctures'. When analyzing relationship between institutions and behavior, the interaction between institutions and individual behavior can be viewed in two ways according to historical

institutionalism. Two approaches have been distinguished within this branch of neo-institutionalism, a calculus and a cultural approach. According to the calculus approach, institutions are shaped according to a sort of 'Nash equilibrium' maximizing each individual's benefit from strategic calculation. According to the cultural approach, people behave strategically albeit bounded by their worldview, which includes moral and cognitive templates provided by institutions. Institutions are therefore resistant to reform because they reflect the very choices about reform that the individual is likely to make. (Hall & Taylor 1996: 5-10). In summary it seems fair to say that the calculus approach assumes that individuals shape an optimal organization that fits their preferences while the cultural approach assumes that an organization shapes individual preferences to fit its cognitive and moral templates.

Sociological (or organizational theory) institutionalism
Sociological Institutionalism began as a subfield of organization theory in late 1970s. It challenged the distinction between rationality and culture. Institutional forms and procedures are largely adopted because they are culturally specific (referring to myths and ceremonies) and not necessarily because they were most efficient. For example: departments of education show international similarities and companies producing comparable products have similarities because of diffusion of techniques and similarities in training. Institutions are defined in terms of symbol systems, cognitive script and moral templates. This determines how situations are recognized and responded to. A division between institutional explanations based on organizational structure and cultural explanations is challenged since the elements that make up a culture (symbol systems, cognitive scripts, moral templates) themselves define an institution.

Comparing the three neo-institutionalisms
Looking at some similarities and differences between these three branches of neo institutionalism, it can be said that both rational choice and sociological approaches suggest that people are purposive, goal oriented and rational. Sociological institutionalism stresses that rational action itself is socially constructed. Rational Choice institutionalism explains the existence of institutions by the assumption that they provide efficiency (a logic of consequence) while sociological institutionalism claims institutions are there because they provide legitimacy (a logic of appropriateness). The central notion from both historical and sociological institutionalism is that new institutions arise amidst existing ones. Sociological institutionalism claims that new institutions borrow templates from existing ones (Bell 2002:8) while historical institutionalism stresses that a new institution is faced with set of existing power relations and path dependencies (Hall&Taylor 1996: 21).

Historical institutionalism is the broadest variant of the three incorporating both a calculus and a cultural approach. Sometimes historical institutionalism is referred to as eclectic (Immergut 1998: 27). In table 2 some key differences between the three branches of neo

institutionalism are summarized. This table may simplify or magnify the complex differences between the neo-institutional branches somewhat for reasons of clarity.

Table 3.2   Key differences between the three branches of neo institutionalism was based on Hall&Taylor (1996:17-21), Immergut (1998:18) and Bell (2002:5-8)

'School' of Neo institutionalism (dominant logic)	*Assumption actors*	*Preferences vs. Behavior*	*Existence of Institutions*	*Phenomena studied*	*Methodology*
Rational Choice (logic of consequence)	Rational, selfish, utility maximizing	Suboptimal equilibriums chosen for strategic reasons	Exist because they provide efficiency	Transaction costs, Property rights, Rent-seeking.	Deduction form abstracted principles
Historical (logic of consequence and logic of appropriateness)	Self reflective	Interpretation of interests shaped by institutional history	Persistent in time, shaping history	Asymmetries of power, Path dependency, Unintended consequences, Inefficiencies of existing institutions	Induction from empirical observations
Sociological (logic of appropriateness)	Cognitively bounded	Unaware of interests because of bounded rationality	Exist because they provide legitimacy	Symbol systems, Cognitive script and Moral templates	Induction from empirical observations

When looking at Performance budgeting and the neo institutionalist branches, it is clear that PB shares many assumptions with Rational Choice Institutionalism. More specifically, PB's popularity can be explained by the fact that it is seen as a solution to a classical Rational Choice dilemma.

3.6   Rational Choice Institutionalism and Agency Theory

It has been noted that rational choice institutionalism developed at the same time as historical institutionalism but finds itself in relative isolation from it (Hall&Taylor 1996:10). Rational choice institutionalists in political science drew fruitful analytical tools from the 'new economics of organization' which emphasizes the importance of property rights, rent-seeking, and transactions costs to the operation and development of institutions. Agency theory, that focuses on the institutional mechanisms whereby 'principals' can monitor and enforce compliance on their 'agents' proved useful for analysis from a rational choice

institutionalist point of view (Hall&Tayor 1996:11). As the solutions that PB suggest particularly feed into the problems highlighted by this branch of neo-institutionalism, it can be argued PB draws heavily on the logic of consequence. By integrating planned activities and expected outcomes and budgets, the existence of a predictable causality in policy theories is explicitly assumed. The expectation that this information will be used in the oversight of government by parliament and in managing government agencies assumes that politicians and managers are to a great extent driven by the expected consequences of policy choices. Furthermore, the need to produce and report PI as a necessary precondition for managing government institutions and their relationships is an essential part of substantial theory building on the part of Rational Choice theorists. A well-known application of Rational Choice theory has been in Agency Theory.

Agency theory and information asymmetry
Agency theory deals with dividing decision making authority within organizations. It takes notice of information asymmetry and the behavioural consequences this can have. Agency theory is built around the relationship between two players in a hierarchical setting; an agent acting on behalf of a principal. To do so, decision making authority and resources are transferred from the principal to the agent. In reality principals are often represented by several agents and vice-versa. An agent representing a principal can itself be a principal to another agent. That way, chains of principal-agent relationship exist. An agent acting on behalf of several principals may face internal discrepancy between conflicting goals and conflicting incentives for behaviour.

If both principal and agent try to maximize their utility, the agent is expected to not fully serve the principal's interests as it will have competing preferences. In agency theory the agent is the expert in performing the tasks that the principal orders it to do. The principal operates at a certain distance from the vital processes and relations, necessary for performing these tasks. As a result the agent has an information advantage that it may use to maximize its own utility at the expense of that of the principal. This so called information asymmetry is the central problem in agency theory (Niskanen 1973: 16-17, Ter Bogt 1997: p.53-55).

Two special cases of misusing information asymmetry are adverse selection and moral hazard. Adverse selection refers to a situation where the principal is presented incorrect or incomplete information by the agent. Given his inability to verify this information the principal may base his decisions on inadequate information. An example is an employer who relies on incorrect résumés presented by jobseekers. Moral hazard refers to a situation where the agent, who has a better awareness of risks and consequences, takes inappropriate risks in the knowledge that the principal will pay for their consequences. An example is a painter who is paid by the hour for a job and therefore decides to take it slowly.

In both of these cases the interests of the principal compete with and are being sacrificed for conflicting interests of the agent. In order to align the agent's interests with his own, the principal can use incentives and monitoring. In addition the agent can engage in activities that help convince the principal that it is acting according to his interest. The expenses made for this are referred to as bonding activities. Agency costs consist of both monitoring costs and bonding costs. The difference that still exists between the actual behaviour of the agent and maximizing utility for the principal, despite making monitoring and bonding costs, is referred to as residual loss.

Bilateral monopoly
In the public sector where the incentives of earning a profit and competition for survival are often absent, information asymmetry can be a major source of inefficiency. Rational Choice assumptions helped inspire some key critical notions on the effectiveness of government bureaucracies. These gained worldwide popularity by way of reforms in governance and financial public management from the 1980s onwards. The assumption of self-maximizing individuals gave way to assumptions of maximizations of budget and staff as primary explanation for the behavior of actors within government bureaucracies. Especially for agency theory, this notion has important consequences. According to principal agent theory, conflicting goals or preferences are a source of inefficiency. The principal's management of its relation with an agent will be aimed at aligning the agent's goals with its own. In doing so, the principal will try to install incentives to steer the behavior of the agent to serve its purposes. Another strategy can be to monitor the agent intensively. These strategies can take the form of direct financial rewards or punishment and varying the degree of autonomy.

Although not considered part of neo-institutionalism, Rational Choice applied to public sector institutions follows similar assumptions as agency theory. Bill Niskanen (1971:24) speaks of a bilateral monopoly to describe the relationship between a public sector bureau (agent) and its sponsor (principal). The agent depends on the financial support of the principal and the principal depends on the agents as the sole provider of particular public goods and services. Each will try to advance its own interest. Sometimes a public sector bureau will advance its own interest without serving any public interest and thus increase the costs of public goods and services (rent seeking). Niskanen expects the relationship between a public sector bureau and its sponsor to result in overproduction of goods and services or overstaffing on the part of public sector agents because of the latter's information advantage (Niskanen 1971:39-41, Ter Bogt 1997: 56-60). The notion of a bilateral monopoly takes the problem of information asymmetry a step further. Not only can a principal expect an agent to take advantage of his superior knowledge to advance his own interests instead of the principal's. The principal's options to do something about this are limited due to the fact that the agent is operating as a monopolist. As a solution, Niskanen suggests organizing competition to the monopolist in an attempt to create some market

conditions and incentives. In the absence of market competition, the principal is left to rely on financial rewards or punishment and gathering critical information by monitoring. Given the importance of information flows for power relations to rational choice theorists, it is not surprising that they engaged heavily in the debate over information asymmetry.

PB as a solution for information asymmetry
The idea of measuring and reporting performance, a core element of PB and PM, has been linked to public choice arguments to control bureaucracies and to principal-agent theory. By making organizational performance data more widely accessible, the information asymmetry between principal and agent is expected to be broken down (Walker et al 2013: 834). The worldwide decentralization of government tasks to (semi) independent agents in the 1980s and 1990s, fed the urgency felt by academics and governments to deal with the principal-agent dilemma. The key question was how ministers can be enabled to better control self-serving bureaucrats. The popular answer borrowed from the market economy was to use contracts that specified the agent's production levels, performance standards and prices. By suggesting these kinds of solutions, applied rational choice institutionalism has had significant impact on the major reforms in public sector management in recent years (Bell 2002: 11). Implementation of performance based contracts in the public sector was usually introduced as an element of broader performance budgeting (PB) and matched the adoption of performance standards for public services. These reforms can be seen as part of a broader package of NPM inspired reforms that include agentification itself (see Section 1.2).

By translating political agenda's into performance based policy plans and contracts with agencies, PB has become a tool to align the interests of political principals with those of agencies. An extensive performance monitoring structure was put in place to bridge the information gap of principals. Sometimes direct financial incentives are being tied to meeting performance targets in attempt to influence behavior of the agency and its members. If working as intended, the agency and its employees are expected to meet the principal's targets in the most efficient and effective way. The expected behavioral effects from information asymmetry (rent seeking, adverse selection and moral hazard) are expected to be offset or prevented with the help of PB. It is worth noting that several authors have doubted the potential of performance information and performance incentives to reduce information asymmetry due to the complexity of many public services and unpredictability of agency responses. (Holmstrom&Milgrom 1991:50, Heckman et al 1997:393-394, Dixit 2002: 697). Moreover, given the numerous occurrences of perverse use of PI by agencies (see CH 1.3) one may argue that, if anything, PB has smoothed this problem at most.

3.7   Theoretical Model
In Chapter Two the multitude of promises and expectations associated with PB was narrowed down to its contribution to operational efficiency and internal result orientation

through purposeful PI use as manifested in the interaction between a political principal and an agency or within an agency itself.

The previous paragraph explained that agency theory highlights the possibility of information asymmetry to frustrate the principal's effectiveness and efficiency by an agent pursuing its own interests instead. According to Rational Choice's assumptions about agency behavior, this is even the general expectation. In the tradition of rational choice institutionalism, the tendency of dysfunctional behavior by the agent has to be offset by other incentives that fit the same logic of consequence. These are financial incentives to perform and monitoring of the agent's performance by the principal. PB has been the weapon of choice by public sector budgeters worldwide to accomplish these tasks over the last couple of decades. If used successfully, PB would help the agent use its resources to achieve the results requested by the principal in the most efficient manner. This way PB is supposed to effectively offset the tendency of the agent to use resources for its own preferences and thus inefficiently or dysfunctional in the eyes of the principal. This reasoning consequently follows to the logic of consequence.

Although there is anecdotal evidence that supports the idea that PB indeed contributes to efficiency this way, applying the principal-agent model to the public sector is considered problematic by some. In his description of the principal agent dilemma Wilson explains the difficulties of applying this model and its assumptions in the public sector (Wilson 1989:156):

The difficulties of avoiding shirking in a government agency go well beyond the problems of meeting the prescriptions of the standard principal-agent theory. First the output of an agency may not only be unobservable, it may be unknowable. If the agency's goal is so vague as to be meaningless (for example: 'advancing the interests of the United States') the administrator often will not know what to do and thus cannot be expected to tell a subordinate what to do, much less judge the work after the fact. Second, every agent in a government bureau is likely to have many principals, not only a bureau supervisor, but also superiors in the Office of Management and Budget, the White House, the courts, and several congressional committees. Moreover the superiors will frequently change as elections come and go. Third, the agents will bring their own political preference, professional standards and prior experiences to their job. .. bureaucrats have preferences and these include definitions of how the job ought to be done as well as how much it ought to pay. For every manager who complains that an employee is doing too little (shirking), there is another one who complains that an employee is doing the wrong thing (subverting).

This leads Wilson to conclude that:
Under these circumstances of vague or conflicting goals, multiple principals, and bureaucrats with policy preferences, it is not surprising that economists have not made much progress in

finding theoretical solutions to the problem of shirking. What is surprising is that bureaucrats work at all rather than shirk at every opportunity.

If the principal agent model and the tools to offset its presumed effects, namely PB, indeed fail to explain behavior in public sector agencies, an alternative way may be required to explain why some organizations and their members act in a result oriented way and use PI for this purpose.

As shown earlier in this chapter, the logic of appropriateness assumes that organizations and organization members are primarily rule followers driven by social pressure instead of anticipating consequences in order to maximizing their own utility. This may offer an alternative explanation of a claim to success of PB. In fact, some recent studies suggest that PI use in the public sector is more likely to be driven by altruism and public sector motivation rather than self- interest among government officials (Moynihan 2010:14). The unpredictable volatile reality of political life in which an agency and its political principal operate may yet offer other additional explanations that may have to be considered. These contextual explanatory factors are factors beyond the institutional sphere of the agency itself. In Figure 3.2, the theoretical model of this research as explained above is shown graphically. This is followed by a brief explanation of each of its consisting elements.

Figure 3.2   Theoretical model of PB functioning according to logic of consequence and alternative explanations
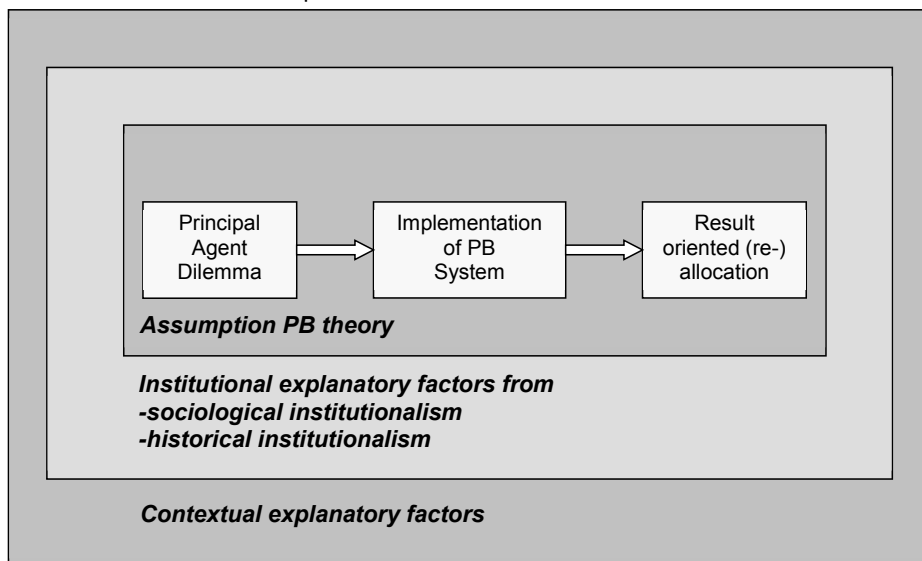
Principal Agent Dilemma

Principal-agent theory is concerned with ways to get an agent to behave in the interest of a principal. The central dilemma in principal agent theory is information asymmetry that occurs when the agent has an information advantage over his principal. The agent thus has the possibility to serve its own interests at the expense of those of the principal. According to public choice theorists the agent can be expected to behave this way, leading agencies to behave inefficiently. The PB intends to help solve this problem by purposeful use of PI by the principal and the agent.

Implementation PB System
PB systems can be categorized according to the degree to which PI is linked to funding (OECD 2007: 21). In some cases a direct link between performance results and resource allocation and accountability is in place (direct/formula performance budgeting). More often the link is indirect and planned performance targets and results are used for planning and accountability purposes only (performance informed budgeting). PB systems that have no link between PI and funding and PI is used for accountability only (presentational performance budgeting).

In government, adoption of formal tools like internal contracts systems or even pay-for-performance does not necessarily mean that these play a serious role in decision making because the relational aspects are often valued more highly than the formal aspect (Schick 2003: 88-91). Implementation of a PB can be merely *de jure* where PI is known by a small number of staff only and seldom shared and debated. When PB is implemented *de facto* managers actually use PI to control their agency or the relations between agent and principal and program learning occurs by critically assessing performance to learn and improve.

Result oriented (re-)allocation
Result oriented (re)allocation refers to budgetary decisions that are linked to the budget cycle and are consistent with results from a systematic performance measurement and reporting. These decisions result in proposed adjustments in an existing program and/or corresponding budget that increase efficiency and/or effectiveness from the viewpoint of the principal. Result oriented (re-)allocation can refer to input, output or outcome criteria, or all of them, as long as it falls within of the formal agreements between principal and agent.

Presented so far is the traditional theory on PB helping to achieve result oriented (re)allocation by an agency. The expected causality between the three elements mentioned above is in accordance with rational choice institutionalism and the logic of consequence. This presumed causality can be challenged by taking into account some of the institutionalist notions presented earlier. As Allen Schick noted: *successful organizations learn and adapt, changing what they do and how they work in response to both internal and external signals. But performance is only one of the drivers of change, and not always the most important one*

(Schick 2003: 88). Two groups of alternative explanations of result orientated allocation will be studied.

Institutional explanatory factors
Several institutional factors may explain why a certain agency displays result oriented (re)allocation. In an attempt to contrast the logic of consequence that has been dominant in the PB approach, elements in other institutionalist schools of thought are identified that are able to explain a possible logic of appropriateness for purposeful use of performance information.

The first two mentioned are motivated by sociological neo institutionalism:
Cultural appropriateness: in a given agency, displaying result orientation behavior by conducting critical assessments or debating effectiveness may be supported or punished by the organization's leadership and culture.
Cognitive frames: The way (performance) information is selected, interpreted, presented and processed may be highly dependent upon shared cognitive frames that exist in an agency. These frames may have originated from a shared organizational or professional or educational background. Implementation of the PB idea of introducing accountability based on objective PI is likely to be heavily influenced by shared cognitive frames.

Another two elements are studied more by historical neo institutionalism, notably the cultural approach within this form of institutionalism:

Asymmetries of power: Some groups or interests have disproportionate influence on decision making process. This formal or informal balance of power may be inherent to a certain institutional setting that characterizes a policy field or an agency.
Path dependency: A critical junction may be identified that created a branching point that explains the current path of the agency in many different respects including de facto PB implementation and result orientated allocation.

Contextual explanatory factors
The dynamics of the policy process can be an explanatory factor itself when the dominant policy paradigm is shifting (as described in Chapter 1.3). In this case enlightenment may occur simultaneously with, though not necessarily as a result of the availability of performance information. Likely alternative explanations are shifts in political leadership or preferences of political leaders. When a public agency is performing politically sensitive tasks or tasks with a high impact, *de facto* political control may be strong regardless the organizational form of the steering relationship designed for agency management (Egeberg and Trondal 2009:6, Pollitt et al 2004 in Verhoest et al 2011: 23). This explains why there may be other explanations for choices of (re)allocation than the incentives provided by PB.

These will be included in the study insofar they provide plausible alternative explanations for an apparent display of result oriented (re)allocation.

<u>Macro-economic factors and shifting political preferences</u> may influence the allocation of resources. The occurrence of financial crises can be a powerful driver of policy changes affecting the principal-agent relationship. The pressure from these factors may lead principal and agent to (temporarily) behave according to PB expectations of result oriented (re)allocation.

3.8   Epilogue

In this chapter we saw that the classical notion that institutions matter to political outcomes and the conviction that political behavior can be explained by observed behavior rather than expressed preferences inspired the rise of neo institutionalism. Depending on the neo-institutional branch, behavior of actors can be explained either by following rules about what is felt to be appropriate (historical/ sociological) or by calculation to maximize self-interest (rational choice). The rational choice branch of neo-institutionalism highlighted the problem that a principal lacks adequate information to control its agent.

The wave of New Public Management in the 1980s and 1990s (see CH. 1.2) reshaped governance relationships in governments worldwide. Performance measurement and evaluation was promoted as one of the tools of choice for implementing the NPM agenda and making government achieve better results at lower costs (Gore 1995: 44-46). Firstly this resulted in adopting public sector management methods and buzzwords like explicit performance standards, management by results, value for money etcetera. Second was an attempt to introduce incentive structures into public service. This meant disaggregating existing bureaucracies and contracting out to quasi markets (Rhodes 1997: 48-49). Performance reporting was presented as part of the solution for the principal-agent dilemma and has been advocated worldwide since the era of New Public Management. As we saw in Chapters 1 and 2, the results of performance budgeting (PB), including this key element, has far from met all expectations. PB theory itself has been accused of neglecting some important institutional variables when designed and implemented. It may therefore be worthwhile to turn to the other two branches of new institutionalism (historical and sociological) when looking for explanations for the success record of PB.

These alternative explanations based on the logic of appropriateness may or may not provide plausible explanations why result oriented (re)allocation occurred in selected cases. This may support or disregard the notion that PB adoption is a likely decisive contributing factor to result oriented (re)allocation. In order to be able to test these cases, the relevant neo institutional factor of the model introduced will have to be refined so they can be applied to the cases.

CHAPTER 4   BUILDING A MODEL TO BE TESTED

Having gained more of a micro-perspective on PB theory, Schick's observation that '*governments that don't manage for results will not budget for results*' (see CH 1.4) deserves a second look (Schick 2003: 102). If it is true what Schick says, introduction of PB by an agency or its principal(s) would only deliver the desired results if a number of other, independent factors, are present. As a consequence, PB successes that are reported from an agency or their principal(s) would, according to Schick, have been made possible only because of these factors that were already in place. He refers to these as '*managerial arrangements that make results paramount*' or '*transformations in public management that enhanced performance*' that accompanied PB reforms (Schick 2003: 102). Expanding on this observation, the question remains what the added value of PB has been in those public organizations that (already) did manage for results. In Chapter 3 neo-institutional factors were introduced that might offer an alternative explanation for purposeful PI use by such agencies. In this chapter the theoretical concepts and research questions will be operationalized into a model with indicators to assess purposeful PI use and institutional explanatory factors. Finally a methodological approach will be chosen and cases will be selected to apply this model.

4.1   Revisiting the research questions

In Chapter 1.4 the general research question of this study was introduced:

Are result orientation of a government agency and operational efficiency gains achieved through PB?

After introducing the micro model of PB, this question was specified further to focus on apparently successful cases of adopting PB as a tool to manage agency-principal relationships and for internal agency management.

How do underlying cultural and historical factors explain successful PB in government agencies?

Chapter 3 introduced the concepts of (neo) institutionalism in which two alternative drivers of human behaviour can be contrasted:

- People act in order to maximize their benefits based on expected consequences (following the *logic of consequence*)
  or
- People act out of duty or social pressure following rules and roles that are they feel are appropriate (following the *logic of appropriateness*)

In Chapter 3 it was argued that PB in its roots, design and implementation dominantly follows the logic of consequence. In recent years an increasing number of authors have stressed that for successful PB implementation institutional and cultural variables outweigh

the importance of the incentives provided by the PB system (see CH.2). If the observation that PB implementation in itself does not change behaviour is correct, it should be possible to explain successful PB implementation using alternative explanations. Instead of PB successfully providing an incentive structure aligning the organization's formal goals with the benefit maximizing behaviour of its members, the displayed behaviour may have entirely different reasons. The result oriented behaviour associated with successful PB implementation may provide legitimacy or be viewed as appropriate by organization members. More specifically an explanation will be sought in four concepts from other branches of neo-institutionalism:

Historical neo institutionalism incorporates both the logic of consequence and the logic of appropriateness (see CH 3.4). The historical branch views behaviour as being influenced by past choices that shape both the perceptions and strategies of actors. Lending from cultural more than a calculus approach, the two key concepts chosen are:
  - Path dependency
  - Asymmetries of power
Sociological neo-institutionalism essentially broadens institutions to the whole range of informal cultural forces[10] (Linsenmann et al 2007: 20-21). As one's interests are shaped by the institutional context in which one operates, the very idea of pursuing rational interests to maximize one's benefits is challenged. Relying firmly on the logic of appropriateness, the two key concepts chosen are:
  - Cultural appropriateness
  - Cognitive frames

A causal relationship between the adoption of PB and occurrence of its desired outcome (result oriented (re)allocation) should be doubted if these factors are present and offer a plausible explanation for the result oriented behaviour displayed in the cases to be studied.

General sub questions
As mentioned in Chapter 2.4, several sub questions are to be to be answered regarding specific cases of successful PB implementation in a principal-agency relationship. By answering these questions the underlying question of how the use of PI might have been regarded as appropriate for the agency, is being addressed.
  1. What is considered a case of successful PB in a government agency?
  2. What are considered relevant underlying historical and cultural factors?
       a) From historical neo institutionalism?
       b) From sociological neo institutionalism?

---

[10] As a result, according to the sociological branch, institutions can be defined differently than traditional boundaries of organization structures. When studying the behavior of agencies, it will be taken into account that characteristics found may apply to institutions that overlap organizational boundaries.

After a number of cases that share (seemingly?) successful PB implementation in an agency-principal relationship, have been studied, the general question can be addressed:

3. Does the presence of these cultural and historical factors offer a convincing alternative explanation for successful PB?

To be able to answer sub question 3, the general questions have to be applied more specifically to each of the case studies so a synthesis of all cases can be created later on.

Case specific sub questions
A   To what extent can a case be considered a textbook example of PB implementation?
B   To what extent are the identified conditions present?
C   Does context offer a likely explanation?
Sub question C explores whether purposeful PI use demonstrated in a case can be explained by other dominant factors instead of the neo-institutional ones tested. This can obviously not be done with the same level of detail. Nonetheless, answering this question gives some assurance that obvious contextual explanations aren't ignored and can be included in the answer to sub question 3. See Figure 4.1 for he order and repetitiveness of the subquestions.

Figure 4.1   Chronological order of sub questions

Sub questions derived from central question:
1: What is considered a case of successful PB in a government agencies?
2: What are considered relevant underlying historical and cultural factors?

Questions for each case study:

A. To what extent is this a PB success?

B Are the identified conditions present?

C. Does context offer a likely explanation?
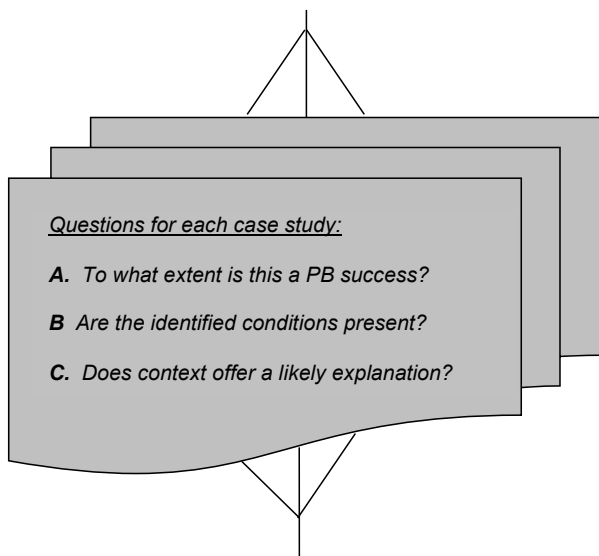
3. Does the presence of these cultural and historical factors offer a convincing alternative explanation for successful PB?

4.2 Agency Theory and Rational Choice Neo Institutionalism

Agency theory states that conflicting goals of agents and principals are a source of inefficiency. As discussed in Chapter 3.6, the agent has the upper hand in this relationship because of his superior knowledge about the particular business in which it operates. The agent will try to exploit this position, potentially leading to adverse selection and moral hazard. PB is introduced as a solution for this problem because performance measurement and reporting is expected to improve the information position of the principal. In addition, financially rewarding performance in budgeting decisions will create incentives for aligning the agent's behaviour with the principal's goals. In the search for potential cases, the likeliness of conflicting goals is assumed and the superiority of the information position of the agent is included in the analysis of the cases. After these assumptions of agency theory are found to be (potentially) present, the problem that PB is supposed to solve can be expected to exist in that particular case. Clearly this is a precondition for any case in this study. Because this study attempts to determine the actual contribution of PB methods on apparent PB success, cases are sought that could qualify as successful cases of PB implementation. Therefore the first sub-question has to be answered:

Sub question 1: *What is considered a case of successful PB in a government agency?*

After answering this question in general, for each case study an answer has to be determined to the associated **sub question A:** *To what extent is this a PB success?*

In Chapter 2 the concept of purposeful PI use was introduced and translated into a model that classified ways in which PI could be expected to contribute to efficiency and effectiveness. Only if PI is used in a purposeful way and thus can be expected to contribute to operational efficiency and internal result orientation the way PB envisaged (see tables 2.3. and 2.4), a suitable case exists for trying to explain PB's actual contribution. This PB ideal can be contrasted to the more 'traditional' way budget preparation takes place within an agency. This was articulated well by Aaron Wildavsky (Wildavsky 1992:90-94). In deciding 'how much to ask for' agency staff takes into account various sources of information, none of them explicitly referring to measured performance:

- *Estimates of what 'will go'.* As the agency is usually faced with more demand for funding than is available, it will ask just below what will be considered 'too much' by the other players in the budgetary process. This way it remains credible when requesting funds for priorities in the future.
- *Informal reports on how its programs were regarded,* especially last year's
- *The letter from the budget bureau,* which usually has some statement on how closely this year's budget should resemble last year's.
- *The interest of specialized publics in particular programs.* Periodic reports from the field on the demand for services may act as an indicator. When the agency begins to

notice connections between the activities of supporting interests and calls from parliament, it has a pretty good idea of the support for a program.

- *Interest of (influential) politicians in particular programs*, mainly 'likes' and 'dislikes' of influential politicians.

The 'traditional' way of budgeting fits the notion of public choice theory that agencies (or public sector bureaus), as a default setting, attempt to 'overproduce' and maximize their budgets. Only the way the budgetary process is shaped and competition from other agencies for scarce resources restrains an even higher bid for funding. According to the theory explained in CH 3.6 and 3.7, effective PB implementation is supposed to be one way to counter the budget maximizing incentives of an agency by handing a principal an additional tool to control its agent. This means that information on future or past performance will be used by the principal to lever funding requests.

Any agency is expected to demonstrate 'traditional' budgeting behaviour to some extent. Neither does the PB ideal described here imply that PI can ever be the sole determinant of budgetary decisions. However, if an agency is making a budget request that refers to PI in addition to the traditional factors, PB will be claimed to have had influence. This goes in particular if the budget request is performance informed and largely disregards 'traditional' determining factors. If PI is used by the principal during budget negotiations, this particularly points to influence of PB.

Since we are seeking to identify purposeful use of PI from a PB system in a principal-agency setting, we can identify between use of PI in the interaction between principal and agent and internal use of PI within an agency. Summarizing the specifications made so far, a 'textbook example' of PB implementation in line with the assumptions of NPM and agency theory is assumed combine four characteristics:

- *PI is used in the budgetary process between agency and principal in addition to traditional budgeting*
- *An intense performance dialogue exists between agency and principal*
- *The agency has an ambitious and sophisticated PB system in place*
- *PI from this system used internally by the agency for budgeting and performance management*

If these four characteristics are found to be present in a public sector agency and the interaction with its principal, PI is indeed used for budgetary purposes and may be expected to positively influence effectiveness and efficiency. Therefore these four characteristics will be developed further into indicators to form a model with the purpose of testing their presence in qualitative case studies.

*PI is used in addition to 'traditional' budgeting (**indicator PB 1.1**)*

According to agency theory and NPM, effective PB implementation is supposed to be one way to counter the budget maximizing incentives of an agency. A principal can do so during budget preparation by using information on future or past performance to lever an agency's funding requests. This requires that performance and funding levels are integrated for the purpose of internal budgetary allocation and for budget requests to the principal. More specifically, PI use in a budgetary context can leads to more efficiency and effectiveness if:

- ✓ PI explicitly informs proposals prepared for the budgetary process;
- ✓ These proposals contribute to program learning or even enlightenment;
- ✓ The proposals result in decisions that demonstrate more efficient service delivery or even policy shifts. This should be recorded in documents related to the budgeting cycle or the principal-agent relationship.

As noted earlier, any government agency is expected to demonstrate 'traditional' budgeting behavior to some extent. Neither does the PB ideal described here imply that PI can ever be the sole determinant of budgetary decisions. PB is however regarded to be applied, if an agency is making a budget request that refers to PI in addition to the traditional factors. This is the case even more so if budget requests are performance informed and if PI is used by the principal during budget negotiations.

*PI use by principal to control the agent (**indicator PB 1.2**)*

In addition to being used in budgetary decisions, PI offers a principal ways to help reduce information asymmetry and thus strengthening its relative position. Performance monitoring or even financial incentives are tools the principal may use to influence agency behaviour in a broader sense than the annual budgetary process. The principal's performance monitoring and its performance dialogue with an agency may vary in nature and intensity. This will influence the principal's ability to successfully use PB to strengthen its position in relation to the agent. PI use is expected to be a major influence in the relation of the agency with its principal if:

- ✓ There is clear alignment of the information monitored and the principal's formal goals.
- ✓ Monitoring data are used frequently in a dialogue between the agency and its principal.
- ✓ This dialogue results in concrete consequences and actions.
- ✓ There are provisions for financial incentives related to performance
- ✓ These provisions are actually used when targets aren't met.

*A high degree of de jure PB implementation (**indicator PB 2.1**)*

To assess the *de jure* implementation of a PB system, some questions regarding its level of sophistication and ambition apply. Although a plethora of theories and opinions exists about

what should make up a good PB system, a categorization that is often referred to internationally is the one introduced in CH 1.1 (OECD 2007: 21, Congiano et al 2013:229):

- ✓ Direct or formula performance budgeting. In some cases a direct link between performance results and resource allocation and accountability is in place.
- ✓ Performance informed budgeting. More often the link is indirect and planned performance targets and results are used for planning and accountability purposes only.
- ✓ Presentational performance budgeting. PB systems that have no link between PI and funding and PI is used for accountability only.

Regardless the appropriateness of the type of PB system in particular circumstances, direct performance budgeting is seen as the most ambitious of these three forms and presentational as the least ambitious one. To further specify the level of sophistication, two additional relevant aspects of *de jure* implementation can be added:

- ✓ What is the frequency of performance measurement and reporting? This may occur only when the yearly budget or report comes around or as part of a monitoring cycle with a higher frequency.
- ✓ Are there measures taken to ensure the quality and impartiality of the reported performance information? Examples are organizational checks and balances or incidental or systematic auditing.

A high degree of *de jure* implementation incorporates an ambitious form of PB with a monitoring cycle more frequent than the budget cycle and measures to ensure the quality and impartiality of information.

*A high degree of de facto PB implementation (**indicator PB 2.2**)*
Having a PB system that is designed in an ambitious and sophisticated manner does not guarantee that PI is used in an internal cycle of performance planning and monitoring and improvement. To measure *de facto* PB implementation, the generation of PI and its subsequent use are important indicators for the way the system works in practice. A high degree of de facto PB implementation is characterized by:

- ✓ Reporting of the required PI in a complete and timely manner.
- ✓ Implicit reference to PI and PB terminology in communication.
- ✓ Explicit reference to PI in communication and decision making.
- ✓ Decision-making matches formal conclusions and recommendations based on expected or demonstrated performance.
- ✓ Budgetary consequences are linked to performance based decisions.

Having defined more closely what successful adoption of a PB system would look like in the setting of a public sector agency and its principal, we are interested to see if our definition matches the situation in public sector agencies that have a reputation of successful PPB adoption and PI use.

The PB result of staff motivation in alignment with agency goals will be considered a favourable by-effect of the other PB effects such as program learning and enlightenment and are expected to influence this aspect of result orientation an indirect way (see Chapter 2.3).

4.3   Historical Neo Institutionalism

This branch of Neo Institutionalism stresses the current influence of changes that occurred over time and, more in particular, the path shaping influence of precedent setting events (Linsenmann et al 2007: 20-21). Historical institutionalism studies asymmetries of power, path dependency, unintended consequences and inefficiencies of existing institutions. For the purpose of this study Historical Neo Institutionalism is used to find an alternative explanation for behaviour that is considered rational and result oriented. More specifically the sub question to be answered is:

Sub question 2a:
What are considered relevant underlying factors from historical neo-institutionalism?

After answering this question in general, for each case study an answer has to be determined to the associated **Sub question B:** *Are the identified conditions present?*

Two notions introduced in Chapter 3 define an institution's *'heritage'*. This heritage is built upon by past decisions that shape current events. This might offer an alternative explanation for the result orientation and purposeful use of PI, attributed to PB introduction.

Path dependency**:** A critical juncture may be identified that created a branching point that explains the current path of the agency in many different respects including de facto PB implementation and result orientated allocation. The presence and impact of such events will therefore be analyzed.
Asymmetries of power: Some groups or interests have dominant influence on the decision making process. This formal or informal balance of power may be inherent to a certain institutional setting that characterizes a policy field or agency.

Path dependency
Historical institutionalists divide the flow of historical events into periods of continuity, punctuated by 'critical junctures'. These are moments when substantial institutional change takes place, creating a 'branching point' from which historical developments moves onto a new path. The problem is to identify the underlying reason for this sudden change. Historical

institutionalists traditionally look for causes like the economy or wars (Hall & Taylor 1996: 342). For the purpose of this study, the occurrence and the impact of such a rupture is, in itself, more relevant that its precise cause. After all an attempt is made to look for alternative explanations for the phenomenon of result oriented (re)allocation by an institution. The question is if there has been a branching point that explains the current path of the agency in many different respects including performance management and result orientated allocation. This branching point can be in the policy field in which the agency operates as well as the organizational arrangement in which it operates.

A critical juncture in the accountability chain or policy field **(Indicator H 1.1 & H1.2)**
As such, when embarrassments, scandals, or disasters occur, politicians and the media suddenly take an enormously detailed interest in organizational activities they have never asked about before. This interest includes performance data (Pollitt 2006:15). Therefore such events may provide valid explanations for an emphasis on PI use by organization members. For example it is quite imaginable that a scandal involving waste of funds or an attack in the media over the inability to deliver expected results can strengthen a culture of results accountability that may already have been present in a weaker form or only locally in certain organizational units. Any type of event may qualify as a plausible explanation of an agency's attitude towards PB but a distinction is made between two broad classifications:

Junctures in accountability chain - One way is by looking at the chain of principal-agent relationships relevant to the agency in question. This chain will be referred to as the accountability chain. A critical juncture somewhere up or down this chain might explain a need to adopt an instrument like PB or even make PB adoption seem unavoidable. This could have been anything but notably causes to look for may be severe financial or accountability problems or traumatic conflicts with political stakeholders about incidents involving performance reporting. If something like this happened, using PB is seen as a solution to a widely felt problem.

Junctures in policy field - Critical junctures can also occur in the policy field in which the agency operates. As noted in CH 1.2, public policy, once implemented, is commonly seen as incremental. If policies are revised in the public arena, this tends to happen in short intervals of turbulence amidst longtime periods of calm (True et al 1999: 176-180). Changed attitudes on the paradigm that traditionally dominated a policy area, might explain the adoption of PB. Again, these junctures may have been about anything but in particular reliance on evidence based policies or the need for collecting previously unrecorded policy information.

Given this interpretation of path dependency it may appear as something hard to measure applied to a particular case. On a nominal scale however this comes down to finding the answer to two questions that can be answered with either a yes or a no (the answer obviously needing convincing argumentation):

68

- Has there been a critical juncture in the accountability chain or policy field?
- If yes, has this critical juncture lead to a broadly felt problem to which PB was seen as a solution.

To investigate the possible contribution of the impact of events on PI use two indicators were chosen:
- The occurrence of a critical juncture in the accountability chain or policy field (H 1.1)
- Did a juncture lead to a broadly felt problem to which PB was seen as a solution (H 1.2)

To assess the relevance of critical junctures in explaining the use of performance information, respondents are asked to name the most significant events that have affected their organization throughout the years. To prevent answers too close to one's own specific duties respondents were asked to think of answers that they felt any colleague might give as well. Subsequently the events are ranked according to the relative importance and the frequency with which they were mentioned by respondents. In addition, the plausibility of their relevance as a possible explanation is assessed.

Some authors stress the importance of keeping path dependence and critical junctures distinct as it is claimed that the former lays the foundations for the latter (Pierson 2000: 263, Pan 2014: 261). At the same time however the former can be punctuated by the latter which inevitable makes such a distinction somewhat arbitrary. As this study focuses on an agency's result orientation and accountability, the criterion for qualifying an event as a critical juncture depends on whether the event can be linked in a plausible way to a major organizational shift in this area.

In line with the sociological institutional tradition Mahler asserts that in a cultural based learning approach, the impact of events cannot be easily assumed in term of lessons for an organization. This is so because the meaning of an event to an organization, like for example a well-publicized failure, is shaped by organizational beliefs and myths (Mahler 1997, 528). However, the events themselves may also contribute to altering or strengthening the very organizational culture with regard to learning and PI use.

Asymmetries of Power
Not to be confused with the information asymmetry central to agency theory, asymmetry of power refers to some groups or interests having dominant influence on the decision making process. This formal or informal balance of power may be inherent to a certain institutional setting that characterizes a policy field or agency.

*An internal advocate or external champion of PB in a powerful position **(Indicator H 2.1)***
Leadership is acknowledges by several authors to positively influence PI use, most notably a leader's commitment to results and his or her support for performance management (Dull

2009a:258-261, Moynihan&Pandey 2010: 14). When the actors pushing PB reforms are in a crucial position within an agency, their agenda's will be more likely to be dominant resulting in higher *de facto* PB implementation. When they are in relatively weak position, PB implementation will be more likely to be halfhearted or *de jure* at most. The actors actively supporting PB have traditionally been budget departments supported externally by the principal's Budget Office. Occasionally these may also have been line managers from the strategic leadership, political leadership or external consultants. To determine the influence of PB advocates, the relative power of the budget office as well as the presence of dominant champions of PB in the agency (or up its chain of accountability) need to be explored. A finding that a dominant party has been successfully pushing PB reforms does not discredit causality between purposeful PI use and PB implementation. It does however provide credibility that, from a logic of appropriateness point of view, PB compliance provides legitimacy to organization members.

*Policy field in which specialists are dominant, relative to the other political actors **(Indicator H 2.2)***
As policy fields are shaped by different groups or interests, the influence of a certain group may dominate the others. A way to analyze the way new issues reach the political agenda and take policy form, is the four fold typology of 'pathways to power' (Posner 2004: 8-14, Conlan et al 2002:4). This typology distinguishes between four different pathways by reference to two dimensions: the scope and scale of mobilization (whether specialized or mass) and the method of mobilization (whether interests or ideas were at play):

- Pluralist pathway (scope=specialized, method=interest): policymaking is driven principally by the process of adjustment among contending interest groups
- Partisan pathway (scope=mass, method=interest): policymaking is characterized by major involvement by political leaders
- Expert pathway (scope=specialized, method=idea): professional knowledge and technical feasibility become the source of legitimacy
- Symbolic pathway (scope=mass, method=idea): ideas and values of broad public appeal play a growing role in defining and legitimizing issues and actions

As earlier mentioned in Chapter 1.3, presenting PI to inform decision making is typically associated with the expert pathway. This specialist perspective on policy competes with the other three dominant ways in which issues reach the political agenda. Although originally applied to political agenda setting, this framework can be used to assess the relative position of the expert pathway in a certain policy area. A policy area in which serving a special interest or providing symbolism for political leaders dominates decisions will be less likely to systematically include PI for decision-making. Once again, a powerful role of the specialist pathway in a policy area does not discredit causality between purposeful PI use and PB implementation. It does however provide credibility to the notion that, from a logic of appropriateness point of view, PB compliance provides legitimacy to organization members.

Although not widely acknowledged as a relevant factor for PI use in literature, this indicator does bear some resemblance to task specificity of leaders which, as opposed to generalist leaders, have been found to report more PI use (Moynihan&Pandey 2010:13). In addition, a high degree of goal clarity and mission orientation is also known to positively correlate with PI use (Moynihan&Landuyt 2009: 1101) and is commonly at odds with policy areas that are highly politicized or dominated by vested interests. Political conflict, as in conflicting interests of internal and external stakeholders, has been hypothesized as a factor limiting PI use in earlier studies but no strong correlation was found (Moynihan and Lavertu 2012: 595, Dull 2009a: 264). The quantitative analysis in both studies did not confirm a strong correlation either. Interestingly, political conflict was also found to foster PI use in another study (Moynihan&Hawes 2012: 101).

4.4   Sociological neo institutionalism

As discussed in Chapter 3.5 this branch of neo Institutionalism challenges the distinction between rationality and culture. Institutions are defined in terms of symbol systems, cognitive script and moral templates. While not denying the value to focus on preferences and interests, sociological neo institutionalism argues that rationality is bounded. The choices actors make are pre-structured by ideas, norms and beliefs about what is legitimate, appropriate or true given a certain conception of identity (Linsenmann et al 2007: 20-21). Sociological Neo Institutionalism is used for the purpose of this study to find an alternative explanation for behaviour that is considered rational and result oriented. More specifically the sub question to be answered is

Sub question 2b:
What are considered relevant underlying factors from sociological neo-institutionalism?

After answering this question in general, for each case study an answer has to be determined to the associated **Sub question B:** *Are the identified conditions present?*

A favorable culture with regard to result orientation in a particular case may offer an alternative explanation for purposeful PI use than the incentives provided by PB. Sociological institutionalism clearly refers to cultural aspects pointed to in PB studies as it stresses the existence and persistence of moral and cognitive templates resulting in the unwritten rules about what is viewed as appropriate behavior by members of organizations. Cultural aspects have been regarded as an important factor for PB success but necessarily make for a somewhat vague variable in research. Two notions introduced in Chapter 3 define an institution's culture incorporating ideas, norms and beliefs that may or may not conflict with the PB system introduced. These are:

<u>Cultural appropriateness:</u> in a given agency, displaying result orientation behavior by conducting critical assessments or debating effectiveness may be supported or punished by the organization's leadership and culture.

<u>Cognitive frames:</u> The way (performance) information is selected, interpreted, presented and processed may be highly dependent upon shared cognitive frames that exist in an agency. These frames may have originated from a shared organizational or professional or educational background. Implementation of the PB idea of introducing accountability based on objective PI is likely to be heavily influenced by shared cognitive frames.

Cultural appropriateness

One clear element is the notion that in order to use PI for enlightenment and program learning, its culture is expected to resemble that of a learning organization. According to some of the more common literature on learning organizations this implies that internal political games that dominate many organizations need to be overcome as they often result in a dysfunctional culture with respect to organizational learning. Typical for this dysfunctional culture is that the question *who* issues a proposal is considered to be more relevant than *what* is in it.

Absence of organizational learning disabilities (Indicator S 1.1)

Peter Senge specified a number of 'traditional' learning disabilities in this respect (see Textbox 4.1). To measure the effect of organizational learning disabilities the response of organization members to these two statements is expected to provide an important clue:

- If (the organization) is faced with problems, a good problem analysis is usually made before taking action.
- If things go wrong in (the organization)'s field of work, lessons are usually learned as a result.

In addition respondents will be questioned about permanent or temporary learning forums that existed and were asked to name examples of lessons that were learned from this for their line of work. The choice to combine cultural and structural approaches to assessing learning organizations acknowledges their mutual interdependency which seems inevitable when studying real life cases in depth (Moynihan&Landuyt 2009: 1097).

Textbox 4.1   Seven learning disabilities according to Senge (source: Senge 1990:18-25)

1. I am my own position is when people focus only on their position within the organization and have little sense of responsibility for the results produced when all positions interact. For example: 'I work as a budget advisor for the health department' instead of 'We try to control public health care costs'

2. The enemy is out there syndrome is when we blame some external reason without focus only on our position; we do not see how our own actions extend beyond the boundary of that position. For example 'politics is just an irrational business' without adding 'which I did not anticipate when making a policy proposal'

3. The illusion of taking charge is that pro-activeness is really reactiveness in disguise. Instead of facing up to difficult issues and solving them before they grow into crises, people rely on more of the trusted ineffective recipes. Illustrative is the attitude 'What we need is a bigger hammer'. Real pro-activeness includes insight in your own contribution to the problem. This is a recognized phenomenon in public policy (see law of accumulation of policy CH 1.3).

4. The fixation on events leads to "event" explanations that are true for now but distract us from seeing the longer-term patterns of change behind the events and understanding the causes of the patterns to events. Practitioners will easily recognize the dominant pattern of political leaders quickly responding to incidents by promising additional regulation and inspection before any deeper, long term analysis has taken place.

5. The parable of the boiled frog is in relation to the maladaptation of organizations to recognize gradually building threats to survival; just as the frog placed in a pot of water brought to boiling temperature will not attempt to jump out of the pot but adjusts to the temperature and slowly dies. The lesson of this story is that we are better equipped to respond to imminent threats than to incremental ones.

6. The delusion of learning from experience is when our actions have consequences in the distant future or are part of the larger operating system, which makes it impossible to learn from direct experience. The most powerful learning impulses stem from direct experience (e.g. tasting). An important notion in this respect is one's *learning horizon*. The learning horizon is the limitation in time and space for judging one's effectiveness. In time this is usually one or two years. Beyond that horizon lessons will have less of an impact. The *learning paradox* refers to the conclusion that our most important decisions are beyond our learning horizon.

7. The myth of the management team because management teams tend to spend their time fighting for turf, avoiding anything that will make them look bad personally, and pretending that everyone is behind the team's collective strategy. Solving urgent problems is usually regarded higher than asking difficult questions on current policies. This poses a problem for organizational learning from performance flaws

*Participative Openness and Reflective Openness **(Indicators 1.2 & 1.3)***
A specific challenge to a learning culture may be deliberate denial or avoidance of certain information to avoid cognitive dissonance. An organizational culture characterized by hiding conflict and protecting management from criticism can block detection and correction of errors (Mahler 1997, 529). This can lead to 'defensive routines' which refers to the policies or actions we put in place to prevent ourselves and our organizations from experiencing embarrassment or threat. The unintended consequence of these defensive routines is that they also prevent anyone from identifying and thereby reducing the causes of the embarrassment or threat (Argyris 1994: 81). Yet failure (particularly in the public sector) is

usually punished – and severely. Thus when a failure is revealed (or even presumed), people tend to hide the deviate data (Behn 2003: 597). The ability of an organization to critically self-assess its performance is therefore an additional important cultural variable for the explanation of purposeful PI use. Senge identified two types of openness that are required for a beneficial culture for organizational learning (Senge 1990: 276-286). These were both tested as separate indicators by asking respondents to react to both associated statements:

Participative openness: Important issues are being discussed openly and fairly
Reflective openness: The ability to continually challenge one's own thinking

Whether this openness exists in an agency is determined by examining both by collecting the opinions of organization members and assessing the examples they present. The analysis of documented meetings can shed some additional light on the attention paid to performance problems and other learning opportunities.

Indicators S 1.1, S 1.2 and S1.3 combined were used as a proxy for cultural appropriateness of PI use. This set of indicators bears resemblance to the feature of a developmental organizational culture and the existence of learning routines by supervisors, two organizational characteristics that have both been associated positively with PI use (Moynihan&Pandey 2010: 14, Moynihan&Landuyt 2009: 1101, Moynihan&Kroll 2015: 22).

Shared view of the meaning of measured performance (Indicator S 2.1)
Another element that links culture to purposeful PI use is the notion that having a dominant single culture or sense of mission can be beneficial to the flow of information within government agencies as described by Wilson:

> *Every organization has a culture, many have several. When a single culture is broadly shared and warmly endorsed it is a mission. The great advantage of a mission is that it permits the head of the agency to be more confident that operators will act in particular ways that the head would have acted had he or she been in their shoes. There are fewer distortions in the flow of information because sender and recipient of the message share common understandings*. (Wilson 1989:109)

As noted by Moynihan the way PI is selected, interpreted, presented and processed may be highly dependent upon shared cognitive frames that characterize an agency as an institution (see CH 2.5). These frames may have originated from a shared organizational, professional or educational background. Implementation of the PB idea of introducing accountability based on objective PI is likely to be heavily influenced by shared cognitive frames. This indicator relates to some of the factors associated with high use of PI in public organizations such as inclusion of organizational members in performance management processes (Melkers&Willoughby 2004: 95) and the ability to link performance measures to one's own actions (Moynihan&Lavertu 2012: 599-600). The same goes, albeit perhaps more indirectly,

74

for public service motivation and the motivational nature of tasks (Perry&Hondeghem 2008: 8, Moynihan&Pandey 2010: 11, Moynihan&Lavertu 2012: 599-600) as a large degree of similarity between organizational goals and one's personal beliefs and convictions could well explain purposeful PI use by organization members.

In her study of the influence of organizational culture on learning in public agencies, Mahler asserts that 'culture plays role in organizational learning by filling in gaps in technological understanding with the collective wisdom of the organization' (Mahler 1997, 536). Viewed this way, the dominance of organizational culture can affect the use of PI to learn and improve as it may heavily influence the way (performance) information is selected, interpreted, presented and processed. If there is a joint understanding across organization members about the meaning of performance data and this interpretation is shared by its principal(s) and other stakeholders, such a culture should be expected to be beneficial to PI use.

However, on the same account, existing institutional knowledge can also ´color´ the perception or interpretation of data in such a way that the organization misses out on opportunities for learning Van der Knaap distinguishes between three results from cognitive framing that limit organizational learning from performance evaluation that also bear relevance to the use of PI by agencies in general (Van der Knaap 1997:62):

- Distorting or blinding function of cognitive frames: existing knowledge 'colors' the perception of a phenomenon or causes organization members to totally miss out on a phenomenon. The result is that different people or groups see something different when observing a phenomenon.
- Existing knowledge 'colors' the interpretation of a perceived phenomenon. The result is that although different people/groups perceive the same phenomenon, they give it another meaning.
- In order to avoid cognitive dissonance a need for reflection that is felt, is deliberately denied or avoided. This response can lead to 'defensive routines' as described by Argyris (Argyris 1994:81)

To assess the helpfulness of cognitive frames for explaining purposeful PI use, the official performance measures as reported by the organization are compared to the respondent's own measure for doing a job successfully. While PB ideally provides objective and relevant PI that is aligned with an organization's goals, organization members may have alternative, often undisclosed, measures for success that they apply to their own work. The fit between the formal performance measures that the PB system provides on the one hand, and the convictions of organization members about what is considered successful on the other hand, will provide a clue about the attitude of organization members towards using a formalized performance measurement system. If there is a large fit, the measured performance is easily recognized as being relevant by the organization and is likely to be used for decisions. If

there is a low fit, information is less likely to be used for decisions and collecting and reporting it is even viewed as irrelevant and too time consuming.

It is likely that cultural appropriateness (indicators S1.1 to 1.3) can also help explain existing cognitive frames to some extent. When there is a taboo on reporting failures, results are more likely to be explained as a success or will be framed to stress only the successful elements. Similarly, in case of an open and self-reflective culture, the risk of denial, avoidance of misinterpretation of PI seems less likely. To avoid overlap, cases will only be studied for the occurrence of cognitive frames other than the ones easily associated with low reflective or participative openness.

4.5   Explaining result oriented behavior

The unpredictable volatile reality of political life in which an agency and its principal often operate may yet offer other additional explanations that have to be considered in addition to the neo-institutional ones considered here. For each case the presence of notable explanations beyond the neo-institutional ones will be assessed. If these are present, explaining result oriented behavior with neo-institutional conditions will be insufficient or maybe even invalid. The latter will obviously be the case if the conditions sought do not occur or only do so to a limited degree. This analysis will form the answer to **sub question C:** *Does context offer a likely explanation?*

Contextual explanatory factors are factors beyond the institutional sphere of the agency itself as mentioned earlier in Chapter 3.7. This means that there are other explanations for choices of (re)allocation than the incentives provided by PB. These will be included in the study insofar they provide plausible alternative explanations for an apparent display of result oriented (re)allocation. Two notable ones stand out but are by no means limitative.

- Macro-economic factors may influence the allocation of resources. The occurrence of financial crises can be powerful drivers of policy changes affecting the principal-agent relationship. The pressure from these factors may lead principal and agent to (temporarily) behave according to PB expectations of result oriented (re)allocation.
- Shifting political preferences can also be an explanatory factor. The dominant policy paradigm can shift because of changed political preferences resulting in a change in political leadership. If this occurs suddenly, a changed policy direction may be marketed as lessons learned from performance measurement or evaluation.

After the cases are examined the information gathered will be compared to formulate an answer to the last sub question:

Sub question 3:   *Does the presence of these cultural and historical factors offer a convincing alternative explanation for successful PB?*

The answer to this sub question will give away the most important clue for answering the central question to this research. An alternative explanation is considered convincing if it matches the findings from the cases. To determine this, a number of hypotheses, based on the central research question and its translation into indicators can be considered. The use of hypotheses for qualitative research concerning a limited number of case studies may be somewhat unusual but was seen as useful for clarifying the interpretation of results concerning concepts that by definition already seem have a tendency remain rather vague. With due observance of the limitations in terms of external validity that may apply to a qualitative assessment of a small number of cases studies, alternative institutional explanations for purposeful use of PI, were aggregated below into five hypotheses.

If almost no evidence from case studies is found that supports the presence of the defined explanatory indicators, the alternative explanations from historical and sociological neo institutionalism will have to be rejected. This would mean that purposeful use of PI has little to do with the presence of these indicators and, indeed, the impact of PBB implementation may have been undervalued. At the very least it would imply that the proper alternative explanatory factors were not identified. As this would confirm the explanation of purposeful PI use from rational choice institutionalism this will be named the R-hypothesis. Obviously this can only be the case if PB adoption indeed fits the criteria for a good practice of PB adoption as determined by the indicators.

Hypothesis R: Purposeful use of PI is explained by the logic of consequence

The refined central research question was: *How do underlying cultural and historical factors explain successful PB in government agencies?* As the institutional context is considered a likely factor that could explain why PB does or does not result in purposeful PI use, alternative explanations for PI use were borrowed from two different schools of neo institutionalism. These can theoretically be contradicted with the rational choice tradition in which principal-agent theory and PB can be situated. Alternative institutional explanations for purposeful use of PI, were aggregated below into four hypotheses:

Hypothesis H1:   Purposeful use of PI is explained by path dependency
Hypothesis H2:   Purposeful use of PI is explained by asymmetries of power
Hypothesis S1:   Purposeful use of PI is explained by cultural appropriateness
Hypothesis S2:   Purposeful use of PI is explained by cognitive frames

Hypotheses H1&2 and S1&2 have been proposed as alternative explanations to Hypothesis R. The contradiction between HR and the alternative hypotheses mirrors the observation that government agencies tend to be more beholden to their culture and traditions than to budget allocations (Schick 2014: 20). If almost no evidence from case studies is found that

supports the presence of the institutional factors, all four alternative assumptions have to be rejected. Figure 4.2 provides the theoretical model of this research in a simplified matter.

Figure 4.2   Using institutionalism to provide alternative explanations of PI use
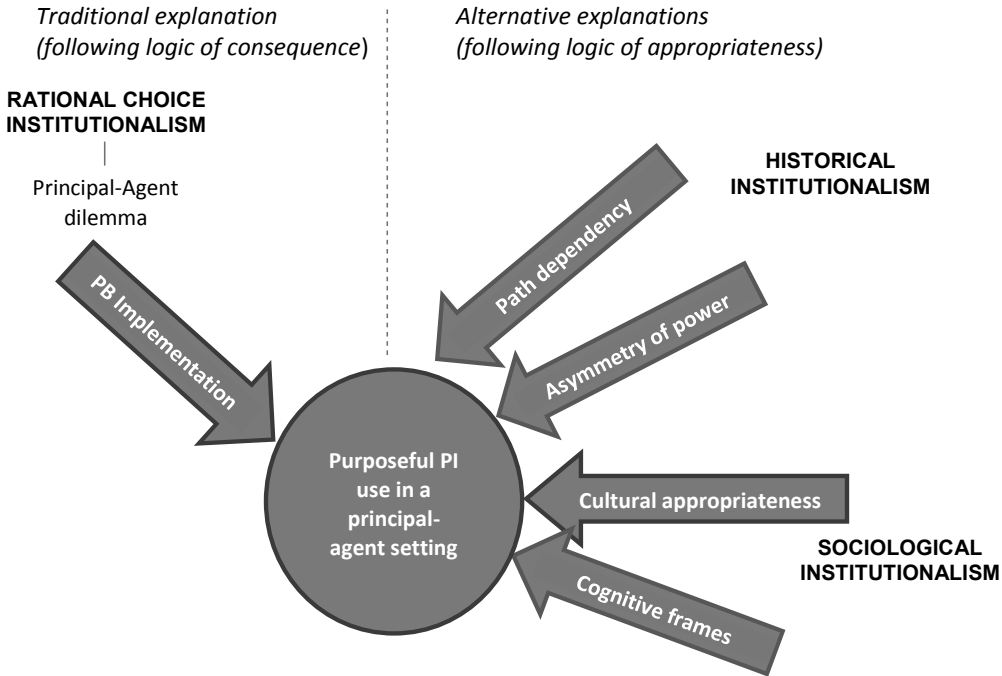
The indicators that were formulated in sections 4.2 to 4.4 can be linked to these hypotheses as can be seen In Table 4.1. It should be added that, given the design of the model tested, the PB indicators are a condition for plausibility of all hypotheses while do other indicators only test the plausibility of the alternative H and S-hypotheses.

Table 4.1   Testing model with theoretical concepts and indicators

Hypothesis and corresponding indicators*	Absent	Present to some extent	Clearly present
Hypothesis R: Purposeful use of PI is explained by the logic of consequence			
PB 1.1	PB is used by the agency in addition to traditional budgeting			
PB 1.2	PB is used by the principal to control the agency			
PB 2.1	A high degree of *de jure* PB implementation			

PB 2.2	A high degree of *de facto* PB implementation			
Hypothesis H1: Purposeful use of PI is explained by path dependency				
H 1.1	Critical juncture in the accountability chain or policy field			
H 1.2	Problem to which PB was seen as a solution			
Hypothesis H2: Purposeful use of PI is explained by asymmetries of power				
H 2.1	An advocate or champion of PB in a powerful position.			
H.2.2	A policy field in which specialists are dominant.			
Hypothesis S1: Purposeful use of PI is explained by cultural appropriateness				
S 1.1	Absence learning disabilities			
S 1.2	Participative openness			
S 1.3	Reflective openness			
Hypothesis S2: Purposeful use of PI is explained by cognitive frames				
S 2.1	Shared view of the meaning of measured performance			

* for further operationalization see the interview questions and the survey in Appendices 1 and 2

If there is strong evidence in support of these alternative explanations, the impact of PB reforms should be seriously questioned. Moreover, the causal relationship between successful PB adoption and the purposeful use of PI that PB adoption was intended to achieve may be reversed. This would imply that favorable institutional circumstances for purposeful PI use, are more likely to explain successful PB adoption than the other way around. If this clearly is the case, using PI may have been driven primarily by the logic of appropriateness. This would provide credibility to the idea that, instead of modifying public organizations into PI users, PB reforms codified patterns of PI use that provide legitimacy to organization members.

Although the H- and S-Hypotheses have been proposed as alternatives to the R-Hypothesis, this does not mean that they are mutually exclusive as a single one or all may be true. It should be noted that the dichotomy between the two types of logics was used here as a theoretical concept to shed some additional light on untested and often unarticulated beliefs on causality associated with NPM reforms and PB in particular. In fact the hypotheses formulated may overlap to some extent which may not be surprising as rational choice and sociological approach are both incorporated in historical institutionalism (Hall&Taylor 1996: 17). Moreover, due to the complexity of causalities investigated, any comprehensive set of

explanations of PI use in a principal-agency setting will likely to interact to some extent. Therefore it is important to realize that the presence of the indicators investigated does not entirely discredit the possibility that PB implementation may have has some positive influence on purposeful use of PI at the same time.

4.6   Methods

To answer the central research question, the assumptions were tested in international cases that share successful PB implementation. This research was designed as a multiple case study because of this method's potential to compare empirical results. The explanations for the phenomenon observed in each of the cases, formed the basis for analytical (rather than statistical) generalization based on the theoretical framework. As diverse cases were selected to explain a particular outcome, namely successful PB adoption, a most different systems design (MDSD) was chosen (source: Blatter&Haverland 2014: 49). In order to prevent that contextual variables would play too big of a role, an adjusted most different system design was chosen. The variety between the cases was restricted by limiting the diversity to two national institutional settings and two policy areas. The assumptions were tested by relying on literal replication from case to case. The choice for literal replication means that a certain degree of similarity between cases is desirable. The study was designed to determine conditions under which the central phenomenon is likely to be found. It does not try to explain contrasting results between cases for reasons that can be predicted form the theoretical model, as in theoretical replication (Yin 2003: 47).

An alternative approach would have been to opt for a Most Similar Systems Design (MSSD) which is used more commonly in comparative studies in public administration. This would have required the selection beforehand of a number of cases that express strong differences with respect to the main independent variable of interest while being as similar as possible with regard to variables associated with other potential explanations (Blatter&Haverland 2014: 25). This was not viewed as feasible as the complexity of the independent variables, as operationalized, practically rule out a careful selection of truly similar cases beforehand. This, combined with, the plural character of the dependent variable (the several aspects of successful PB adoption) made an MSSD approach unlikely to result in clear and reliable conclusions.

Performance based reforms not only enjoyed worldwide popularity, the available evidence also suggests that their impact on performance did not distinguish significantly between national contexts (Gerrish 2015:20). As the phenomenon investigated refers to worldwide experience with result oriented budgeting, cases were sought in different policy fields in different countries. Even though this required additional analysis of relevant differences between countries in terms of political and budgeting systems, it does offer additional insights and prevents that overall conclusions are only associated with one particular policy area in one national context. To limit the number of independent variables, related

programs in terms of policy fields were sought in different countries. Although a larger number of cases or a choice for 4 different policy areas in 4 countries may have increased external validity to some extent, for reasons of limited time and capacity just four case studies were conducted. For the method of case selection see section 4.7 and Appendix IV.

The primary method of data collection were semi structured interviews. This combined the flexibility to obtain unexpected relevant information while ensuring focus on the information required by the analytical framework. Interview candidates were selected to represent management, performance & budget staff and operations of each agency as well as staff tasked with oversight on the part of the principal (see Appendix II for exact functions). In addition several experts and senior government officials [see Appendix II for name and position) were consulted on particular issues. Each indicator was scored by the author on a qualitative 3-way scale ('not present', 'present to some extent' or 'clearly present'). The rating of interview questions was done while re listening to the interview recordings and the result was recorded in an interview form for each interviewee. For some indicators a scale of criteria was used to identify between the 3 categories while in others a choice was made while referring to key remarks from the interviewee. In Appendix II the interview form used for each respondent can be viewed. Data were primarily processed by qualitative analysis although some quantitative analysis was used for comparing and aggregating results. These scores were aggregated for each case allowing for a qualitative comparison on an ordinal scale. After finishing each draft case study, the report was reviewed by a key informant from the organization involved.

A specific risk to validity in interviews was responder bias. Translating this specifically to the subject of this study, it is notable that to many a system of PB is something that is theoretically hard to disagree with. Possible discrepancy between adoption of formal PB system and actual result orientation is specifically targeted in this study. This requires profound knowledge of signs hinting at such discrepancy before conducting interviews. Specific methods that were employed to promote accurate and completeness of newly acquired information from interviews were audio-recording, pre-coding of answers in interviews and recording data shortly after acquiring them. By means of triangulation the findings from interviews were compared with questionnaire results and findings from analysis of documents. For example this enabled a comparison between the descriptions of cultural openness as provided by the interviewees with the anonymous responses from questionnaires. Similarly, different accounts of the actual use of PI for budgetary purposes from documents, interviews and questionnaires could be detected. Appendix III shows the questionnaire that was used.

Gaining access to key organization members was expected to pose a technical difficulty. Using a large professional network and support from the advisory committee was helpful in gaining the required access. However this also poses a risk of being associated with certain

interests or maybe even jeopardizing these unintentionally. Working in the field of budget accountability at the NL Ministry of Finance means that the author is subject to the apparent advantages and risks of being a practitioner/researcher. Apart from the obvious advantages of being a 'pracademic' in the same field (access to data, synergy, practical knowledge and experience) this combination is known to present a couple of risks of its own. These are: professional bias, lack of time and methodological expertise, a professional or hierarchical bias towards research results and the tendency of many organizations to value outside advice more highly (Robson 1993: 534-538). Earlier research experience of the author, a balanced committee in which both academics and practitioners were represented as well as frequently discussing results with colleagues and fellow researchers is expected to have contained these risks.

4.7   Case selection

The choice for the Netherlands and the US federal context results from the author's familiarity with both from earlier research and work as well as the advantage of large accessible data availability that the U.S. government offers in this area. Case selection started with creating a shortlist of potential cases in the Netherlands based on government reports and academic papers as well as talks with academics and experts from the Netherlands Ministry of Finance and the Netherlands National Court of Audit. This resulted in a shortlist of thirteen principal-agent relationships (see Appendix II) in which output based funding and performance management are claimed to have been dominant in recent years. To control for obvious international differences in policy characteristics, consultations with experts in the US and the Netherlands resulted in a selection of five candidates out of thirteen (see Appendix II). These five were selected because the public services they provided were expected to be least affected by national policy context. For these candidates, the US federal counterparts were analyzed with help of the White House's PART database[11]. All of these five sufficiently met the criteria for a suitable case study and although two other policy areas offered slightly superior opportunities on paper, forestry and air traffic control were selected on more pragmatic grounds after discussions with US and Netherlands experts. Factors determining the final case selection were the ability to acquire additional specialist knowledge and the availability of contacts with relevant professionals. Finally the occurrence of an extensive multi-year reorganization in one of the candidates, i.e. NVWA (Food Safety Administration of the Netherlands), discredited it as a practical candidate. The results of the final case selection are presented in Table 4.2. For

---

[11] PART stands for Program Assessment Rating Tool. This tool was used by the Office of Management and Budget during the Bush (jr) Administration. Under PART about 1.000 U.S. Federal programs were systematically rated in terms of performance. For this rating a standardized question list was used that heavily focused on the availability and use of performance information in the management an d budgeting processes.

each case study organizational entities were chosen to represent the central principal and agent. Nonetheless additional principal-agent relationships were inevitably included like the ones between an agent's (regional) headquarter and geographical unit. In Appendix IV more details are provided regarding the case selection process.

Table 4.2   Results case selection

Policy area	NL Principal-Agent	US Principal - Agent
Forestry	**SBB:** Min. van LNV / EL&I – Staatsbosbeheer	**USFS:** USDA/HQ US Forest Service - Pacific Region 9
Aviation (air traffic control)	**LVNL:** Min. van VenW / I&M – Luchtverkeers-leiding Nederland	**FAA/ATO:** Dept. of Transport - FAA/Air Traffic Organization

4.8   Case description

To enable a fair comparison between the cases studied, each will be described according to a similar format. After a general description of the case (paragraph 1) the question will be addressed to what extent the case represents a PB success (paragraph 2). Subsequently explanations provided by the selected neo-institutionalist concepts will be examined (paragraph 3). Then the analysis will turn to the context to see if it offers a likely explanation (paragraph 4). Finally a conclusion will be drawn considering the scores on the indicators and the question which hypothesis is supported by the case (paragraph 5). Each case will end with an epilogue that will look back on the key findings and additional observations as well as the data collection for the case. A more elaborate format for the description of each case is provided in Appendix V.

CHAPTER 5 CASE STUDY STAATSBOSBEHEER
(National Forest Service of the Netherlands)

5.1   Description of the Agency and its Principal

History and tasks SBB
Staatsbosbeheer (SBB) is an agency that manages 260.000 hectares of public lands in the Netherlands on behalf of the national government. Although part of the public sector, some of the activities SBB undertakes also make it a player on private markets. As a hybrid organization, SBB and its political stakeholders have been attempting to balance the execution of public tasks with benefiting from commercial incentives. SBB's main principal has been the Ministerie van Landbouw, Natuurbeheer en Visserij (Agriculture, Nature and Fishery) which was merged into the Ministerie van ELenI (Economic Affairs, Agriculture and Innovation) in 2010. In 2012 the Ministry van ELenI was renamed to Ministerie van Economische Zaken (Economic Affairs).

In 1998 SBB was granted a semi-independent status by law. This was the result of a discussion about possible privatization that had already started in the early 1980s. The objectives of the 1998 law ('Wet Verzelfstandiging Staatsbosbeheer) were twofold. Firstly SBB would be able to provide a better product at lower costs. These lower costs would have to result from efficiency gains, output steering, the introduction of accrual accounting and generating additional own revenues from commercial activities.
Secondly, by gaining more of a separate identity, SBB would be better able to present itself to stakeholders in a recognizable way. This would help SBB to meet stakeholders' wishes and amongst these notably those of citizens[12]. Or, as someone at SBB put it: forestry people will have to learn that people are no longer considered a harmful species in the woods

SBB has long upheld a good reputation with regard to performance management. However, in the face of massive cuts envisioned by the current Dutch government (up to 80% by some accounts), this method of steering is beginning to lose some of its relevance. Therefore in the investigated period, the era before around 2008/2009 is by many accounts a different one that the years before. According to plans of the minority government lead by PM Rutte, the responsibility for the main national tasks that SBB performs on behalf of the national government, was to be transferred to the 12 provinces of the Netherlands.

This case study takes into account the period from the mid-1990s up until now. SBB's history however goes back a lot longer that. Staatsbosbeheer (SBB) was established in 1899 when

---

[12] Schmitz, E.M.A. e.a. 2003:7, Commissie Evaluatie Staatsbosbeheer. *Vooruit op Eigen Benen – Evaluatie van de verzelfstandiging van Staatsbosbeheer 1998-2003* (evaluatie vindt elke 4 jaar plaats uit hoofde van art.32 van de Wet Verzelfstandiging Staatsbosbeheer)

the Netherlands was severely deforested and demand for wood was at a high. Also soil erosion was a major problem. In its early days, wood production remained SBB's main emphasis, regulating wood supply by managing stocks of production forests. In the late 1960s the Netherlands could no longer compete with timber producers in larger countries. Also, a change in thinking about nature and environment led to growing criticism of production forests. From then on, focus altered towards the development of natural, mixed forests.

As early as the beginning of the twentieth century, protection of 'nature' and preserving 'natural monuments' became more important issues. In 1928 conservation officially became SBB's second task. The 1970s saw a more offensive approach to nature in the form of active nature restoration and development.

Two more tasks were added in the early 20th century. From 1915 SBB advised the Department of Public Works on the planting of roads and waterways. This grew into a third task: landscape planning. As a result SBB has had a great deal of influence on the formation of the Dutch landscape for decades. As manager of large parts of the landscape, these days, SBB focuses on the preservation of scenic values, the regional identity of the cultural landscapes and historic elements.

From the 1920s, people began to explore the recreational value of nature. Creating and maintaining recreational facilities and managing the growing flows of visitors became SBB's fourth task. It is SBB's policy to make their sites accessible to all Dutch public and offer a range of opportunities to experience nature.

In response to the social developments, SBB today uses its experience in the management and development of multifunctional nature. Varied use of space requires versatile control that contributes to[13]:
- the quality of living, working and recreation;
- biodiversity;
- the resilience of nature;
- maintaining the unique identity of the Dutch landscape

Size
Headquartered in Driebergen, SBB currently employs around 1000 people. In 2010 SBB's annual turnover totaled € 151.1 million. Around 60% (€89.5 million) was contributed by the Ministry of EL&I while some 30% was collected from own activities (€ 48 million) notably from timber sales which accounted for about half of this amount. The remaining 10% came from small subsidies. These figures have been fairly stable over most of the last decade,

---

[13] www.staatsbosbeheer.nl

however under current austerity measures SBB's contribution from the Ministry will be cut dramatically. In addition, national tasks would be passed down to the provinces. The level of funding to be transferred to the provinces as well as their commitment to these plans remains unclear at the time of writing.

Organization

SBB is divided into 4 regions that are primarily tasked with land management and services such as terrain management, recreation and education. In addition two separate branches work nationwide on specific tasks:

SBB- BuitenZaken (Services) is tasked with service delivery from SBB activities to clients, both consumers and businesses. Beside lumber sales other commercial activities are delivery of biomass, potting soil and rental of cabins and rural houses.

SBB-Grond en Gebouwen (Terrains and Buildings) advises regional SBB management about maintaining about 1700 buildings including cultural and historical buildings located in SBB's terrains.

Figure 5.1   Organizational Chart SBB (source www.staatsbosbeheer.nl)
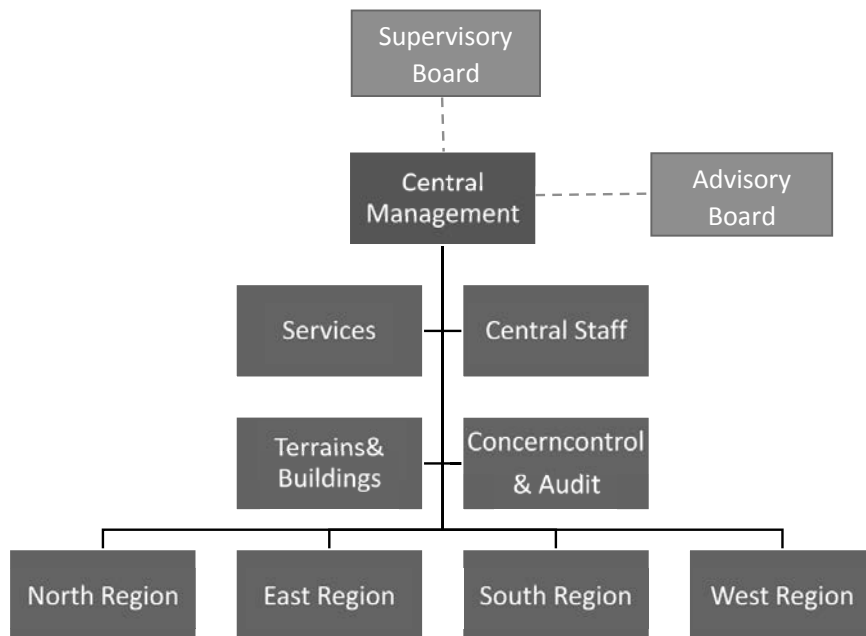
Each of the four regions is organized into several districts under which a number of business units like terrains and visitor centers resort.

Figure 5.2   Organizational Chart of a SBB region (source: www.staatsbosbeheer.nl)
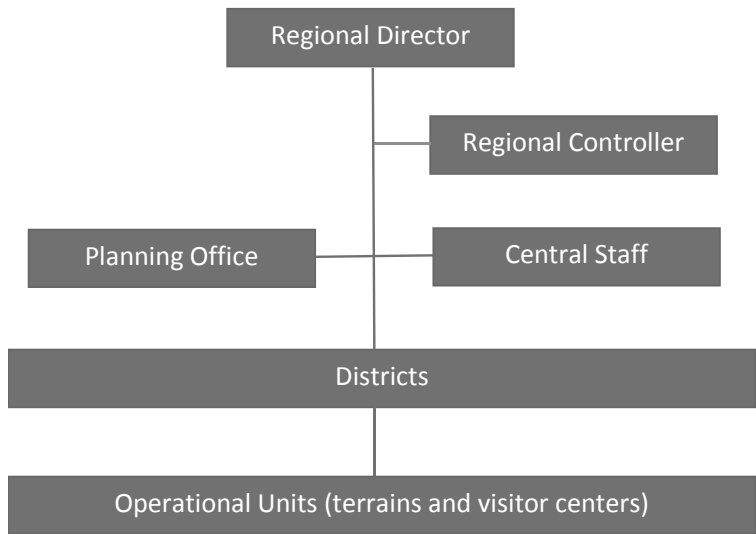
Policy field and tasks
The national tasks that SBB has been performing by on behalf of the Ministry are currently stated under a single program goal in the latter's 2012 budget documents

Article 18.3: Maintaining international and national biodiversity and strengthening our natural resources.

In previous years a larger number of programs and objectives were used. These objectives have been altered numerous times over time. The land management objectives used to be part of the formal objectives in recent budget years but were dropped from 2012 in anticipation of decentralization of these tasks. In order to give an idea what the Ministry's formal objectives looked like in previous years, the policy goals from the 2007 budget are listed in Appendix V.1 as an example of a typical year.

In addition to the funds SBB gets from government, SBB BuitenZaken (Services) is generating revenue from commercial activities. These activities are undertaken with some cautiousness given SBB's hybrid character. In 2012 the legal basis for SBB selling products and services was strengthened following a series of unsuccessful legal actions taken by competitors. Also it is expected that the diminishing relevance of the national government as a stakeholder, will mean less political intervention with initiatives undertaken by this branch of SBB.

5.2 Degree of PB implementation

As mentioned previously, SBB has maintained a certain reputation when it comes to performance budgeting and performance management. This reputation is largely based on its system of output funding. In this paragraph the characteristics of SBB's performance budgeting system are analyzed both in its relationship with the Ministry as internally.

5.2.1 PB in relationship Ministry - SBB

The planning and reporting relationship between SBB and the Ministry fits typical NPM characteristics in the sense that policy execution has been separated from policy development and that SBB is funded on the basis of output targets. Each year SBB's offers a proposal to the Ministry containing an activity program, funding levels and performance indicators. The primary goods and services SBB offers fall into several categories of output:

- Management of terrains for maintaining Nature and Landscape
- Management of terrains for Recreation
- Representation of interests
- Development of new natural areas
- Hosting visitors

The bulk of Ministry's contribution to SBB is intended for terrain-management. The associated outputs are stated in number of hectares with a price per hectare dependent on the type of terrain. In addition some miscellaneous activities and financial arrangements are mentioned in the offer. After negotiation between both parties this results in an order on behalf of the ministry, the offer is made into an order from the ministry which forms the basis for SBB's annual budget. It is worthwhile noting that the principle that the Ministry determines WHAT should be delivered and SBB determines HOW it should be delivered is mostly implemented in ecological targets and lumber production targets[14].

During the budget execution year, high level talks are held twice a year between SBB's director and the State secretary. In spring the previous year's account is discussed while in the autumn, next year's offer is discussed. The Ministry's Director Nature meets with SBB's deputy director about every month. In addition there are meetings of a less formal nature between SBB and lower ranking civil servants that take place about once a month.

The budget articles of the Ministry's annual budget have sported a number of performance indicators on amongst others biodiversity, terrain management outputs, accessibility of terrains and land acquisition. In the Ministries' yearly report the realization of these figures in relation to their target values are reported. In addition SBB provides external stakeholders

---

[14] D'Hondt E.M. e.a. 2009: 78, Commissie Evaluatie Staatsbosbeheer. *Ruimte in het Bos! - Evaluatie Staatsbosbeheer*. 2009 (evaluatie vindt elke 4 jaar plaats uit hoofde van art.32 van de Wet Verzelfstandiging Staatsbosbeheer)

with two accountability documents. One is the financial annual report and the other one is about realization of policy objectives. The latter is quite specific on goal realization. For example the 2010 report stated that 78% of the nature goal types had been realized. Poor planning on the side of SBB was stated as being the cause for 41% of those goal types not realized. This displays a relatively high degree of critical self-reflection in the context of Dutch public administration.

Actual use of PI in relationship ministry-SBB
Despite the frequent contacts, SBB respondents indicated that they seldom got a response from the ministry regarding PI they reported. This is a development that has deteriorated over time and is explained by an increasing lack of specialist knowledge at the ministry. The quotes below illustrate the image SBB has of the ministry's role as a principal:

> *'The ministry is not really interested in any output steering system, their management is mainly political'.*

> *'In fact the Ministry has been granting us a lump sum without any further steering in the last couple of years'*

During its existence as an independent agency, the relationship between SBB and the ministry has been subject to changes over time. Some legal conflicts have occurred regarding lumber sales and a critical appraisal of SBB's role in society, claiming it was still too internally focused. On its part, SBB claimed the ministry often disregarded its specialized knowledge during policy development. Moreover SBB experienced that despite its newly gained status, Parliament and the ministry did still interfere with its business when politically sensitive issues were at stake. During the 2004-2006 period, tension rose about the output based system of funding (price per hectare). This led the ministry to hire an external party to do counter expertise on SBB's offer[15]. In 2007 these conflicts were settled in an adjusted output based system that has been used since between the ministry and SBB. Ironically, upon completion of this system, the major budget cuts and decentralization were announced, undermining full implementation of this new methodology. It is likely however that in the relationship with its future principals, the provinces, the same methodology will be used.

During times of underfunding of desired outputs (occurring from 2006 or so) the ministry and SBB yearly specify work that SBB would attempt to deliver but would not be held accountable for. This was put in annexes to the order. This implies that the output

---

[15] D'Hondt E.M. e.a. 2009: 73, Commissie Evaluatie Staatsbosbeheer. *Ruimte in het Bos! - Evaluatie Staatsbosbeheer*. 2009 (evaluatie vindt elke 4 jaar plaats uit hoofde van art.32 van de Wet Verzelfstandiging Staatsbosbeheer)

management methodology is taken seriously by the ministry albeit in a way rather different or even inverse from the image of the principal controlling the agent.

5.2.2   Performance management within SBB

Ecological targets, required outputs and activities play a pivotal role in the PB's intricate planning system. This system has been adapted over time to fit the ministry's wishes as well as standardization with other (private) terrain managing organizations in the Netherlands. The result is a myriad of plans of different frequencies and levels of detail ranging from long term plans of up to 20 years to annual plans. These plans state terrain characteristics, target types of flora and fauna, and the activities needed to keep or attain the intended target types. During allocation, funding is directly based upon the activities in the plan with occasional adaptation to local circumstances.

Monitoring of terrains takes place continuously by terrain managers. Every 3 months heads of districts and their regional director meet. In addition the SBB director and the regional directors also use quarterly meetings to take stock. On the agenda are mostly execution problems or current events as well as finance. Output information from monitoring is occasionally used in these meetings but not in a systematic matter. The Internal Quality system ensures that each year about 10% of SBB's terrains throughout the Netherlands is scrutinized for goal realization with the help of external auditors. The variety of the terrains selected and the thoroughness of the evaluation results in lessons that are relevant to the entire organization. It is therefore reckoned that for their yearly monitoring, managers can rely on only a limited number of indicators.

Terrain management and ecological goals as well as SBB's commercial activities rely heavily on quantitative output information. For recreational policy this is less the case due to the diversity and volatility of policy objectives in this policy field. Ex post calculation is done only at the macro level at the national headquarter. Discussing accountability of outputs versus costs at a detailed level within districts is viewed as a pitfall to be avoided. This is felt because it takes up much time and the additional insights for steering purposes are limited.

Having looked at the performance budgeting system as formally adopted, perhaps more interesting is the actual role PI plays in decision making. Over 75% of respondents at SBB confirm that they rely on PI in their work (32% regulary and 46% occasionally). Perhaps surprisingly, longer time SBB employees more often claim to be using PI regularly. Regarding the implementation and use of PB methods over time, several phases can be distinguished over the years:

- Development (1980s to early 1990s)

Prior to PB implementation money, people and content were treated as separated silos. Planning was a lengthy and costly process that actually had little relevance.

Exemplary for those days was the following anecdote shared by one of the respondents:

'According to planning, a sand road had to be improved by leveling it with a special device called a land grader. However, each year the available funds were spent on other things then buying a land grader. As a result this action just re-occurred in the planning year after year.'

In the early 1990s a system was developed that was based around certain target types of nature (vegetation and fauna). For every type of landscape, the necessary actions were identified to bring it up to the desired standard to support that particular type of nature. Sometimes not much has to be done to attain a target whereas in other cases this is a costly affair because it requires more 'artificial' measures (e.g. keeping water-levels high or mowing grass regularly). That way an objective basis was created for coupling outputs to activities and costs.

- Initial effects (1998-2004)

The new system worked quite well at the start. Its internal effects were twofold. Firstly it stimulated efficiency by enabling managers and workers to creatively make use of their budget. For example the balance between doing work yourself and external hiring shifted. One respondent remembers:

'Prior the new system, the allocation of funds to districts was a yearly nightmare. Everyone brought on arguments that explained why they would require extra money. Having large numbers of steady personnel was seen as the best strategy for a district to safeguard a large part of the available funding. The objective output measures ended this at once. The terrain characteristics and target types of nature determined the funds needed. This meant that if you had few personnel, you would still get the money you required. This created an incentive to no longer base your number of employees on the yearly peak time but instead on the low-season. In peak time flexible workers were hired temporarily. This resulted in large efficiency gains.'

A second effect was that people became more aware of the intended results of their work. This awareness stimulated creativity that benefited efficiency and effectiveness. With regard to the ability to measure effectiveness, an ecologist tasked with collecting data from field observation recalls that:

'In the 1990s terrain managers lost the autonomy to collect the ecological data they wanted. Up until then this was largely a matter of personal preference. One terrain manager would monitor the number of a certain butterfly species, while the other would count his favorite type of flower. After implementation of the new system, the data

collected had to match the target types of flora and fauna attributed his terrain. This
enabled implementing of a national policy for biodiversity.'

- Perfection the control system (2004-2007)

In their attempts to further refine the system, technical experts overcomplicated the output based system with ever more subtypes of terrains. In the meantime professional controllers tried to use the system for continuous monitoring of targets. The result was that people from the districts and business units put too much trust in the system instead of thinking for themselves. The fun and creativity were slowly wielded out. This was not beneficial to attaining further efficiency gains.

- Adjusted control system and large budget cuts (2007 - present)

The system of output controls as agreed with the ministry in 2007 has been audited twice to check the validity of cost prices. Despite this, the ministry's annuals contribution only covers between 50 and 60% of these cost prices so far and this may be getting even worse. Obviously this may undermine the system of output steering at some point. Right now managers in the districts and terrains use their professional expertise to set their own priorities within the given budget level. Using a more pragmatic approach, the central control philosophy relies upon simplicity and consistence nowadays using PI only when useful to support the internal dialogue and learning processes.

It is important to note that the *SBB BuitenZaken* (the commercial activities branch) has developed its own system of output monitoring without much interference from principals. For recreational activities the number of visitors, number of stays, occupation rates and customer satisfaction are monitored. Actual production of goods (e.g. lumber and soil) are monitored in conjunction with their cost prices and sales proceeds. This monitoring is characterized by a higher frequency than SBB's other activities.

5.2.3 Conclusion on PB implementation

From the point of view of the ministry, quantitative planning and measurement of nature and terrain management is well developed by SBB. As mentioned earlier, more than three out of five employees claim that PI is used in their work. In the fields of nature and recreation this is less the case. Other than *SBB Dienstverlening,* quantitative targets are scarcely mentioned and are hardly used for steering and giving account in these two fields.

Of the different uses of PI mentioned in CH 2.3, all uses were said to occur. Over 95% of the respondents was familiar with the actual use of PIfor at least one of the categories mentioned, with an average of 3.8 uses selected by each respondent. Most often mentioned were the utilization of PI for the purposes of external accountability and internally reallocating resources (see Table 5.1).

Table 5.1   Utilization of PI by SBB according to respondents

Utilization of PI for:	% of respondents choosing this answer in the questionnaire
Performance reporting for external accountability	70,6%
Strategically reallocate internal resources	57,6%
Setting program priorities	37,6%
Allocating internal funds	29,4%
Monitor cost and performance and contract management	24,7%
Analyzing productivity and funding levels	24,7%
Motivate staff to act consistent with goals	21,2%
Deciding on outsourcing decisions	20,0%
Adopting new program approaches following evaluation	16,5%

The five categories printed in bold represent the five most frequently chosen categories given by SBB interviewees. A notable difference is that using PI to adopt new programs was mentioned more often by interviewees. This may be explained by the fact that management and bureau staff were relatively overrepresented in the interviews and are more likely to be involved in policy development.

According to the micro model of PB introduced in Chapter 2, these uses of PI count as successful PB implementation if they contribute to the phenomena of enlightenment and program learning, leading to more efficient and effective ways policy execution. The examples reported from the era of initial effects (1998-2004) within SBB do seem to fit this criterion.

The cyclical use of PI in the different evaluation cycles, most notably in SBB's internal quality system, also hint at program learning. Without much effort an abundance of these types of program lessons can be named by respondents. The value of the evaluation results for financial planning however is not completely transparent. Although the link between ecological objectives (target types of nature) and financial consequences is a rather direct one, it is clear that the primary process is the dominant driver for cyclical performance reporting and evaluation and not the financial planning and control cycle. In other words: it occurs like finance has been linked to output rather than outputs to finance. Looking at the indicators for PB success, the evidence of actual use of PB by the ministry to control SBB is hardly convincing. However the internal effects of performance management and the role of outputs in the internal budget preparation are clearly visible.

Table 5.2   Presence of indicators of performance budgeting implementation

Indicators		Presence in SBB case
PB 1.1	PB is used by the agency in addition to traditional budgeting	Clearly present
PB 1.2	PB is used by the principal to control the agency	Absent
PB 2.1	A high degree of de jure PB implementation	Present to some extent
PB 2.2	A high degree of de facto PB implementation	Clearly present

5.3   Exploring Neo-Institutionalist Explanations

The main question of this research is to explain purposeful PI use by an agency, in this case, SBB. It is beyond doubt that SBB has been using performance measurement to enhance efficiency and effectiveness in performing its public tasks. However that does not automatically mean that this has occurred thanks to implementation of a PB system. Lending from neo institutional theory, several alternative explanations for this result oriented behavior are explored.

5.3.1   Explanations from Historical Neo Institutionalism

When investigating path dependency as an explanatory factor, important events with an impact the organization were sought that might explain PB implementation. This could be anything, for example a crisis with a principal or a series of bad press in the media. Moreover it is possible that, if adoption of PB is viewed as an appropriate response to such events, result oriented behavior should be explained by the logic of appropriateness rather than PB's incentives doing their work. To put it differently: organization members are susceptible to unwritten rules and roles that demand compliance with a performance management system, regardless of the incentives of this system itself.

In the interviews respondents were asked to name the top 3 events that have affected their organization throughout the years and were asked to think of answers that they felt their colleagues might give as well. The latter remark was added to prevent that respondents mentioned events too close to their own specific duties. When comparing the answers mostly mentioned were:
- The shift towards more customer orientation to citizens and stakeholders
- Gaining of independence status in 1998 and, not surprisingly
- The budget cuts and decentralization plans of the current government

13 different answers were given with the most frequently mentioned event having been named by only half of the respondents. This alone does not support the occurrence of an event with the impact to qualify as a critical juncture for SBB. In addition these events most mentioned do not bear a clear connection with the introduction of SBB's system of performance budgeting. At most it can be argued that a system of output funding has been a condition that helped enable the independent status.

Another alternative explanation for adopting a PB system might be the dominance of a certain influential unit or person that acted as a champion of output measurement or performance management. It is quite common that the central financial unit pushes for adoption a PB system or that a particular director is keen on implementing such a system. Although the name of the SBB director at the turn of the century was named a couple of times, no single powerful champion of PB was found. In reality there is more evidence that the initiative for the output based budgeting system was system happened bottom up by specialists from different regions albeit supported by a few key figures at the central headquarters. More in general the position of control-units at SBB does not seem to be dominant relative to management. Their role in developing targets also seems to be relatively limited.

Finally, a dominant role of specialists in a certain policy field may explain why objective measurement is regarded higher than other more political factors. To determine this, respondents were asked to choose between factors that are most influential to SBB policy in two subsequent questions[16] :

A) Vested interests   or   B) Policy ideas
    and
A) Politics/public opinion   or   B) specialist expertise

If for both questions answer B is chosen, this indicates that SBB's policy area is run pretty much by experts and their policy ideas. Most respondents described that up until 2008/9 this was indeed the situation for SBB. Specialist could do their work without much political intervention or even attention. Interests obviously did play a role, mainly during policy execution. It wasn't until a few years ago that the policy paradigm of specialists was discussed at the national political level. This in particular applies to ecological policies. Based the questionnaire, the recent situation is reflected mostly in the answers as illustrated in Table 5.3.

---

[16] Based on 'Pathways to Power' framework, (Posner 2004:8-14)

Table 5.3   Dominant factors driving SBB's policy area

Dominant role in policy area	A) Vested interests	B) Policy ideas
A) Politics / public opinion	31,9% (Partisan pathway)	41,9% (Symbolic pathway)
B) Specialist expertise	11,3% (Pluralist pathway)	14,8% (Expert pathway)

It can be concluded that, unlike the other indicators from this category, the dominance of specialist in SBB's field of policy may offer somewhat of an alternative explanation of the strong emphasis on measured performance. This assessment however may fit yesterday's SBB better than today's.

Table 5.4   Presence of indicators historical neo-institutionalism

Indicators		*Presence in SBB case*
H 1.1	Critical juncture in the accountability chain or policy field	Absent
H 1.2	Problem to which PB was seen as a solution	Absent
H 2.1	An advocate or champion of PB in a powerful position	Absent
H.2.2	A policy field in which specialists are dominant.	Present to some extent

5.3.2   Explanations from Sociological Neo Institutionalism

Sociological neo institutionalism introduced the concept of cultural appropriateness as a possible explanation of behavior, in this case behavior associated with the ideal of PB implementation. In organizations that are characterized by a culture that is open to organizational learning, more favorable conditions for using PI for enlightenment and program learning can be expected. This situation can easily be contrasted with many government organizations in which bad news and self-criticism are met with suspicion and defensive routines. The conditions for organizational learning were tested by letting respondents react to four statements:

    a.   Within SBB, important issues are being discussed openly and fairly

    b.   Within SBB, existing opinions are regularly challenged and discussed

c. If SBB is being confronted with a problem, a thorough problem analysis takes place prior to taking action

d. If things don't work out in SBB's policies and execution, lessons are usually learned

The responses to these questions in the interviews reveal SBB as an organization that is close to resembling an archetype learning organization. When looking at the answers from the questionnaire a somewhat different picture arises:

Figure 5.3   Average of responses to the 4 statements on organizational learning
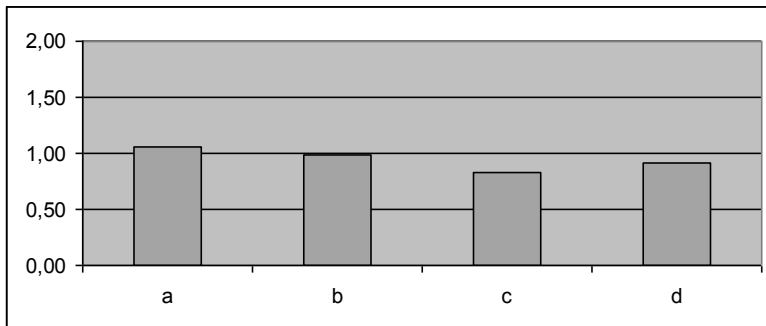
(0 = totally disagree, 0.5 = somewhat disagree, 1 = neutral, 1.5 = somewhat agree, 2 = totally agree)

The average score on these statements contrasts with the responses from the interview. Despite the same guarantees regarding confidentiality, social desirability may have played a role in the responses given in the interviews. In addition the majority of interviewees worked held staff and management positions. It is possible that fostering learning culture is more actively encouraged at (regional) headquarters.

Indeed the headquarters sees an active role for itself in stimulating the sharing of organizational knowledge. The reason for this has to do with the fact that the average forester is usually a dedicated professional who is used to working autonomously. Therefore he will, by himself, not easily visit a colleague to see if he can learn from him. Although well able to self-reflect and challenge existing positions, the commitment to decisions taken earlier is sometimes mentioned as a shortcoming the interviews do reveal that that there are some subjects that are characterized by somewhat more sensitivity or secrecy. Sometimes, sensitivity to outside developments and support for SBB's policies is mentioned as still somewhat of a weak point.

As a result of the idealism and intrinsic motivation that many employees share, the distinction between a person and his professional ideas can be somewhat more of an issue. This can sometimes prove to be an obstacle to learning behavior. Overall though, the willingness to share knowledge and learn seems to be strong according to the interviewees as is the ability to self-reflect. They also agree that these organizational capabilities have

grown in recent years. This view is supported by the existence of several highly regarded learning forums within SBB.

Another important factor that enables the use of a PB system for organizational learning is the absence of distorting cognitive frames. Distorting cognitive frames hamper a learning dialogue because the signals from, in this case performance management, are interpreted differently by members of the organization. This was investigated by asking respondents about their attitude towards performance indicators, their knowledge of SBB's performance measurement and whether they believed it was beneficial to their organization. The answers from the interviews indicated that SBB's performance management system was well known and seen as relevant by most. This indicates that results from performance measurement are quite undisputed and have a similar meaning to different people in the organization. Looking at the questionnaire, the average of the 4 questions measuring attitude towards and knowledge of SBB's performance measurement system, varied according to the position held by the respondents:

Figure 5.4   Attitude towards and knowledge of performance measurement system
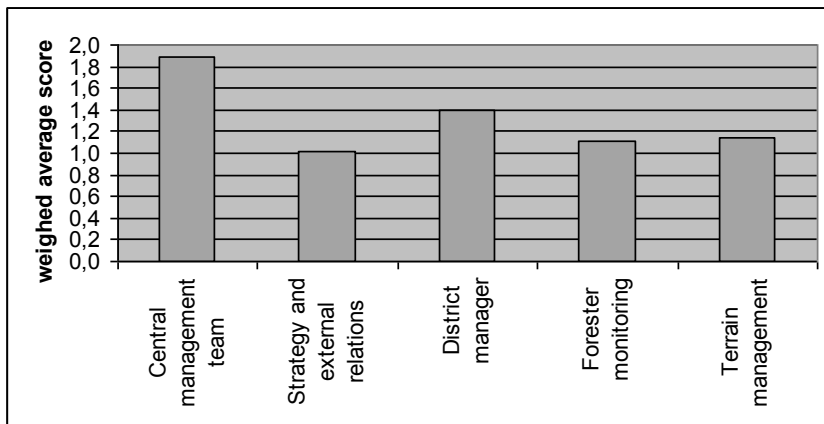
From a viewpoint of sociological neo institutionalism, the ability to openly discuss problems and reflect on existing positions make for an organizational culture in which using of PI for learning and enlightenment is regarded as appropriate behavior. This is especially more likely since SBB's performance measurement itself is viewed as a legitimate assessment of relevant results.

Table 5.3   Presence of indicators sociological neo institutionalism

Indicators		Presence in SBB case
S 1.1	Absence learning disabilities	Present to some extent
S 1.2	Participative openness	Clearly present
S 1.3	Reflective openness	Present to some extent
S 2.1	Shared view of the meaning of measured performance	Clearly present

5.4   Contextual factors

Two contextual factors should not be ignored regarding the SBB case. First of all the impact of the fiscal crisis has landed particularly harsh on SBB leading to budgetary insecurity. In addition the move to decentralize most SBB tasks from the Ministry to the 12 provinces will lead to an increase of the number of primary principals to deal with. So far this has had some impact in the sense that the interest for output reporting from the side of the Ministry was reduced and the internal steering and quality assurance system is under increasing pressure.

Secondly, the shift in political preferences regarding conservation of nature has seen a dramatic change. This was expressed by one of the respondents in a straightforward matter:

> '10 years ago, you didn't have to explain anybody why nature was beneficial for people. In those days SBB developed from a terrain manager to a protector of nature. Nowadays SBB's policies do require explanation as managing terrains can also be done by private parties. I am surprised at the fierceness of these changes in recent years.'

It remains to be seen how permanent and irreversible these developments are for SBB. For all the impact they've had and undoubtedly will have, they do not provide likely explanations for purposeful PI use by SBB as they only occurred relatively recently.

5.5   Conclusion

The investigation of the indicators for SBB should provide answers to three questions to be answered for each case (see CH 4):

A. *To what extent is this a PB success?*

B *Are the identified conditions present?*

C. *Does context offer a likely explanation?*

The answer to the first two questions can be expressed when the numerical scores, obtained from the interviews are expressed in the range of 0-2 and applied to the different indicators:

Figure 5.5   Numerical scores of the indicators from the interviews
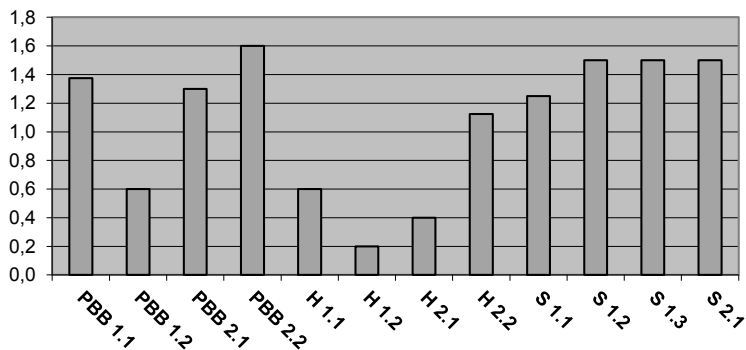
The figure shows that SBB does fit the criteria for a PB success in most respects except for the use of PB by the principal to control its agent (PB 1.2). The score on indicator PB 1.1 was confirmed by the questionnaire that only measured this PB indicator in a direct manner.

The second question can be answered positive for an open culture and the absence of distorting cognitive frames (S 1.2 and S 2.1). Additionally, the relative dominance held traditionally by specialists in PB's policy field (H 2.2) and a culture relatively favorable to organizational learning (S 1.1, S 1.3) are also present albeit less clear as these findings are only partly confirmed by questionnaire results.

In response to the third question it can be argued that in recent years two significant contextual factors have affected SBB. However, during most of the period studied these have been absent.

Summarizing these findings, from a neo-institutionalist point of view, SBB's culture of openly discussing results and a commitment to learn and improve make it appropriate for organization members to actively engage in monitoring, evaluation and fully accepting funding based on outputs. In addition, the relative dominance (at least until recently) of specialists in SBB's field of work did mean that the content of the work was hardly affected by political ideology or mobilization of mass interests. This meant that performance management system was able to develop undisturbed without many questioning its relevance. So it seems that in the SBB case the logic of appropriateness does offers explanations to the result oriented behavior displayed by its members.

Epilogue

The intrinsic motivation of SBB staff may well be the key variable here. Although hard to come by in most ministerial settings, a passion for monitoring the results of one's work perfectly fits any performance management system. It does not seem unlikely that an intrinsically motivated person who is interested in saving a particular endangered animal or restoring a type of landscape, is genuinely interested in the results of policy measurement and opportunities for policy improvement. The rather loose coupling between monitoring results and money except in the planning phase does fit this image. Why would an intrinsically motivated professional be interested in efficiency as long as this does not interfere with attaining his goals?

Finally there is the Nature of SBB's work. It is notable that in the fields of work of SBB, output measurement stems from a long standing tradition. In ecology, the presence and abundance of indicator species are traditionally used as an indication of the well-being of a larger group of species. Or, as one of SBB's ecologists mentioned:

  'Working with indicators is a phenomenon that comes with our profession.'

In the lumber-industry and terrain management, quantitative planning methods have been around long before the modern concept of performance management. The existence of countable work output, quality standards and standardized lists of specific jobs as well as the needed amount of time to perform them, were already noticed by Herbert Kaufman in his classic study of the U.S. Forest Service (Kaufman 1960: 115). For SBB this was illustrated by an anecdote of one of the respondents:

  'While researching the development of lumber prices with a colleague back in the early 1990s, we came across stack of old year plans form the 1930s in a shack in the Veluwe[17]. It occurred to us that these had the same layout as the planning documents we used back then with virtually the same categories of types of terrains, activities and inputs'

The fact that SBB's recreational tasks were the underdeveloped part of its performance management system may have to do with the fact that these activities seem to lack the idealistic or commercial component that seem to be present in SBB's other activities.

Given this background, it is hardly surprising that the system developed by SBB in the 1990s was largely a bottom – up effort by regional staff. Again, this somewhat contrasts the traditional idea of PB being introduced by the principal as a tool to control an agent.

---

[17] Veluwe is the name of the largest national park of the Netherlands, located in the center of the country.

Some have argued that due to the nature of the tasks performed by a Forest Service, 'tendencies toward fragmentation' such as geographic dispersion and varying local conditions pose extra challenges to attaining organizational unity, compliance and conformity (Kaufman 1960: 86-87, Tipple&Wellman 1991: 422). Admitting these factors are physically far greater for the U.S. Forest Service than for SBB, it may equally serve as an explanation of organizational culture of both organizations. Out of necessity, management will have to rely heavily on the autonomy and discretion of foresters to manage their own terrains. In such an organization it is not surprising that devoted and self-reliant (if not obstinate) individuals fare best. Intrinsic motivation and a clear and stable sense of purpose have had to make up for the absence of an intense system of monitoring and incentives by management because this was never a realistic option. If we were looking to define result orientation within public sector organizations, intrinsic motivation, self-reliance and a clear sense of purpose may be key elements.

Although the institutional characteristics of SBB and its policy field may explain result oriented behavior for a great deal, it is not self-evident that new organizational goals will be smoothly adopted throughout the organization incorporated into the daily work of employees. To the contrary, professional autonomy may well prove to be an obstacle in this respect. We have seen that during the phase of 'initial effects' (se par 5.2.2) SBB's PB system, including its financial incentives, may have played a role in goal alignment. So PB may have been successful in helping direct existing result orientation towards new organizational goals more than to actually stimulate result oriented behavior as such. One might even argue that NPM's key premise of results accountability combined with decentralization (or steering instead of rowing if you will) more easily fits an institutional profile of self-reliance and intrinsic motivation.

Data collection SBB case
The primary method of data collection were semi structured interviews. Between December 2011 and February 2012, ten persons were interviewed. Two of these worked at the Ministry of Economic Affairs and were charged with oversight on SBB: one of them focusing on financial oversight, the other one on meeting policy objectives and performance goals. At SBB HQ four persons were interviewed: one representing each of the following units: central management, Services, Central Staff and Concerncontrol & Audit. At the Eastern Region HQ, 3 persons were interviewed: the regional director, the regional controller and a chief-forester (see Appendix II for more details). Data were primarily processed by qualitative analysis although some quantitative analysis was used for comparing and aggregating results (see section 4.6). These scores were aggregated for each case allowing for a qualitative comparison on an ordinal scale. After finishing each draft case study, the report was reviewed by a key informant from the organization involved. By means of triangulation the findings from interviews were compared with anonymous questionnaire results and findings from analysis of documents. For example this enabled a comparison between the

descriptions of cultural openness as provided by the interviewees with the anonymous responses from questionnaires. Similarly, different accounts of the actual use of PI for budgetary purposes from documents, interviews and questionnaires could be detected. Appendix III shows the questionnaire that was used.

Sources

Schmitz, E.M.A. e.a. 2003, Commissie Evaluatie Staatsbosbeheer. *Vooruit op Eigen Benen – Evaluatie van de verzelfstandiging van Staatsbosbeheer 1998-2003* (evaluatie vindt elke 4 jaar plaats uit hoofde van art.32 van de Wet Verzelfstandiging Staatsbosbeheer)

D'Hondt E.M. e.a. 2009, Commissie Evaluatie Staatsbosbeheer. *Ruimte in het Bos! - Evaluatie Staatsbosbeheer*. 2009 (evaluatie vindt elke 4 jaar plaats uit hoofde van art.32 van de Wet Verzelfstandiging Staatsbosbeheer)

SBB - Interne Kwaliteitsbeoordeling 2010, object Oosterwolde

SBB - *Interne Kwaliteitsbeoordeling op terreincondities en doelcomponenten*, eindrapport object Sallandse Heuvelrug 2009

Staatsbosbeheer Offerte 2008 - *Natuur en Jeugd*

Staatsbosbeheer Doelrealisatie 2008 – *prestatieverantwoording aan het ministerie van LNV*

Sturingsarrangement Staatsbosbeheer LNV 2009

Staatbosbeheer Jaarstukken 2010

SBB - Doelrealisatie 2010

SBB - Jaarverslag 2011

Staatsbosbeheer CCA Jaarverslag 2011

Kansen benutten - ondernemingsplan Staatsbosbeheer 2012-2015

SBB - *agenda eerste kwartaalgesprekken regio Oost* april 2011

SBB - *Uitwerkingsplan Sallandse Heuvelrug*

SBB - Achterhoek H1, H2 e.v. (inventarization and planning documents)

Bedrijfscoderingen 'uitgaven' en 'inkomsten' voor terreinbeheer (per 1 jan 1990)

Naar Buiten (SBB Magazine), 14 e Jaargang nummer 4, December 2011

Buitenleven 2012 (SBB catalogue rental cabins)

Erfpacht bij Staatsbosbeheer, SBB folder mei 2011

Pollitt, Christopher et al. 2004, *Agencies – How Governments do Things Through Semi-Autonomous Organizations*. Basingstoke and New York (Palgrave MacMillan)

CHAPTER 6 CASE STUDY LVNL   (Air Traffic Control the Netherlands)

6.1   Description of the Agency and its Principal

History and tasks LVNL
Luchtverkleersleiding Nederland (LVNL) is a agency responsible for managing the majority of
air traffic to and from the airports in the Netherlands. The task of civil air traffic control (ATC)
within the Dutch airspace is shared with the Royal Netherlands Air Force that handles civil
aviation from military airfields and Eurocontrol for the *en route* air traffic that is just over
flying the Netherlands. LVNL handles over 3000 flights daily, almost 90% of which take place
at Amsterdam's Schiphol national airport (based on 2011 figures). In addition to directing air
traffic, other tasks of LVNL include maintaining and replacing technical systems, providing
information to aviation stakeholders, training air traffic controllers and produce aviation
related publications like aeronautical charts.

LVNL operates amidst a number of dominant private, public and semi-public stakeholders.
The most important ones are Schiphol airport, KLM-Air France, the Ministry of Infrastructure
and Environment and the European Commission. The latter two fulfil the role of principal in
the principal-agent relationship regarding LVNL's public task. The EC has started to act as a
principal only in recent years following adoption of its policy in 2004 to bring about a Single
European Sky (SES). As a result the European sky is divided into Functional Airspace Blocks
(FAB). The Netherlands, Belgium, Luxemburg, France, Germany and Switzerland agreed to
form a single FAB in 2010 referred to as Functional Airpsace Block Europe Central (FABEC).
LVNL and its counterparts in the other FABEC member states cooperate to meet
performance targets that were agreed with the European Commission.

The Ministry is chosen as the main principal in this case study because it has been fulfilling
the role of principal during the entire time frame considered from the mid-1990s up until
2012. In addition, the target setting, performance reporting and financial consequences
brought about by the European agreements are integrated entirely in the principal-agent
relationship between the Ministry and LVNL.

The history of air traffic control in the Netherlands dates back to the 1920s. Initially air traffic
control tasks were handled solely by the military. After the establishment of the first civil
airports in the Netherlands in the early 1920s, civil air traffic control started operations in
1923. After an aviation law took effect in 1927, the government established the
Luchtvaartdienst, renamed *Rijksluchtvaartdienst* in 1930. In these days radio was the
primary tool used to direct air traffic. After World War 2 radar was introduced to supplant
the radio. Until 1992 air traffic control in the Netherlands remained a part of the
Rijksluchtvaartdienst (RLD – national aviation authortity) called the *Luchtverkeersbeveiliging*
(LVB). In 1993 the LVB gained a legal independent status within the aviation law. The LVB
was renamed LVNL in 1998. Various reasons are given for granting the status of independent

government agency. Generally the desire for a more modest central government and for increased effectiveness and efficiency in performing public tasks played a role. The idea behind this construction was that this would largely immunize LVB for political interference in their business processes. Another consideration suggested for granting LVB an independent status in 1993 was the incompatibility of salaries of air traffic controllers with the restrictive salary regulations of civil servants (Volkskrant 1999, LVNL 2009). ATC salaries are largely driven by demands from this particular niche of the international labor market and in some cases LVNL's salaries do indeed exceed the standards set for civil servant standards. Internationally however, LVNL does not seem to be out of line in this respect. By some the independent status was considered as a step towards full privatization, an intention that never got enough support to be materialized. Long after LVNL gained its independent status, European regulations, aiming to bring about a Single European Sky (SES), also demanded a functional separation between the air traffic service provider and a supervisory and policy department.

Size
LVNL employs around 900 people approximately 250 of whom are air traffic controllers. The number of employees was brought down from over 1000 following a reorganization in 2009. Annual turnover has been growing steadily each year to € 184,4 million in 2011. This is collected from tariffs paid by airlines that make use of LVNL's services, most notably by KLM-Air France, the major airline in the Netherlands. In terms of costs, about some 75% is made up by personnel expenses and about 10% on depreciation of LVNL's assets.

Organization
LVNL delivers air traffic control services from four locations. The first and by far most important of these is Amsterdam Schiphol airport. The others are located at the airports of Rotterdam-The Hague, Groningen-Eelde and Maastricht-Aachen. Due to the close proximity between Amsterdam and Rotterdam both locations functionally resort under a single unit: *Operations* (Ops). The other two locations resort under the *Regional Unit* (RU). Besides these operational units there are a number of units that support LVNL's primary process in a more or less direct way. The Strategy and Performance (S&P) unit is tasked with strategic planning, performance measurement, reporting and communication. Procedures and Systems (P&S) is responsible for designing and reviewing the air traffic control procedures as well as maintaining the technical systems LVNL employs. Legal Affairs (LA) and Corporate Services (CS) are responsible for more regular staff duties such as legal support, finance and operational management.

LVNL's management team consists of the CEO, the CFO (who jointly form the Executive Board), the directors of OPS and P&S and the managers of CS and S&P.

Figure 6.1   Organizational Chart LVNL
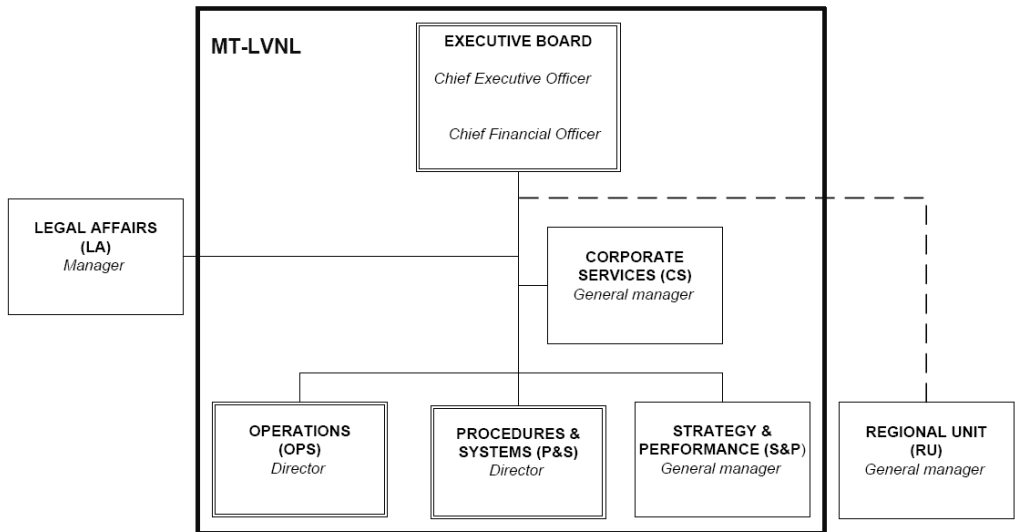
Policy field and tasks

The policy field of civil aviation is characterized by diverse interests of different stakeholders. Although most stakeholders can agree on the desirability to maximize the safety of air travel, doing so may mean compromising on other policy aims like maximize airport capacity, minimizing noise levels or minimizing tariffs for airlines. As a result the policy goals of the national government are also conflicting to a certain extent. This is well reflected by the fact that LVNL's public tasks were stated under different program goals in the Ministry's 2012 budget documents:

Article 33.3   *In order to attain permanent safety improvement in the aviation industry, government sets preconditions for safe aviation operations* - this program goal refers to the activities and funding for policy design and international coordination aimed at reducing the number of incidents

Article 35.1   *Strengthening the competitive position of Schiphol mainport as well as regional airports* - this program goal refers to the efforts to maximize capacity within the agreed noise levels

Article 35.3   *Strengthening connectivity to the international airline network* - this program goal refers to the efforts to prevent delays and to minimize costs for airlines operating from Schiphol

Before a major reshuffling of programs and budget articles from 2006, LVNL's policy field was to be found in a another set of articles. The government's diverse objectives regarding safety, airport capacity and environment were also apparent in these previous budget articles.

6.2   Degree of PB implementation

LVNL is known as one of the agencies that integrated its performance dialogue with the Ministry of Transport with the budgetary cycle. In addition its internal system of performance management stands out as one of the most advanced in the public sector. In this paragraph the characteristics of LVNL's performance budgeting system are analyzed both in its relationship with the Ministry as internally.

6.2.1   PB in relationship Ministry – LVNL

LVNL is funded by tariffs paid by airlines. Its budget however is due to approval by the Ministry of Transport. Determining the level of tariffs is a matter of national policy since it directly affects the competitiveness of the airports of the Netherlands and therefore has significant economic impact. Currently the national airport Schiphol ranks within the top 5 of largest airports in Europe.

LVNL's costs and performance are presented jointly in the annual plans and reports and are both monitored by the ministry. Prior to issuing these plans to the ministry before November 1st, LVNL discusses them with the other major stakeholders. During November budget approval by the Ministry takes place including the tariff levels for the year to come. A set of performance targets and indicators are part of the annual performance plan and are subject of monitoring by the ministry. Each quarter the financial results, the performance objectives and indicators are reported by LVNL to the Ministry and discussed in a bilateral meeting . It is fair to state that LVNL is continuously involved in a rather intense dialogue with the Ministry. In addition to bilateral contacts between LVNL and the ministry, both meet frequently in structured multilateral meetings with LVNL's clients in the aviation industry as well as in specific improvement task groups. The efforts of the Ministry are primarily geared towards balancing the conflicting interests of the different stakeholders like controlling noise levels for citizens, reasonable tariffs for airlines, ensuring enough airport capacity and, of course, ensuring safety for air travellers. Looking at the airline industry as a whole it should be noted that the goals of the ministry and LVNL are perceived to be more aligned with each other than with those of other parties (airports, airlines, citizens) who typically have a more single focused goal.

From 2004 a system of formal performance steering was pioneered in anticipation of European ambitions regarding a Single European Sky. Another reason to further formalize performance steering on the part of the ministry was criticism from the Court of Audit that the ministry was unable to assess LVNL's functioning and performance pointing to

significant risks regarding asymmetry in expertise and information (Algemene Rekenkamer 2004). In accordance with the annual budgetary cycle a large number of performance indicators are reported by LVNL to the ministry, close to 40 in numbers according to the ministry's 2005 evaluation. From 2006 to 2010 an annual 'order ' from the ministry to LVNL (*beleidskaderbrief*) was issued specifying expected performance levels. The set of performance indicators that LVNL reported to the ministry were picked by LVNL from existing measured data. The relation between cost and performance was however mostly limited to presentation. The options the government has when it comes to controlling LVNL and its performance and costs are actually quite limited given the impact of international and commercial stakeholders. However, despite the fact that PI is not directly linked to costs it does play a vital role in the dialogue between the Ministry and LVNL. The performance dialogue became more serious as LVNL succeeded in intertwining its activities, procedures, outputs and outcomes using the so-called VEM methodology (see 6.2.2). Initially this system was meant for internal use but had clear potential for utilization in the dialogue with LVNL's stakeholders (Ter Avest, 2005). Indeed this framework did help to mature the performance dialogue with the Ministry in the sense that LVNL's performance became less of a 'black box' to the Ministry and the other stakeholders,

From 2006, concerns about noise levels as a result of growth of Schiphol national airport saw the birth of a semi-permanent advisory board to the government. The so called Alders Conference *(Alders tafels* named after its chairman) work out detailed agreements about permitted noise levels. LVNL is an important party in these agreements because it has the knowledge to oversee the consequences for safety and airport capacity of proposals meant to limit noise to people living in the vicinity of Schiphol. Furthermore LVNL is expected to do its part in complying with these agreements by taking into account the noise impact when selecting the proper runway (of which Schiphol has five). The ministry's Inspectorate is in charge of monitoring LVNL's compliance with these politically sensitive agreements.

In addition to reporting to the Ministry of Transport of the Netherlands, a more traditional PB relationship including financial incentives has been formed with the European Commission. This so called FABEC agreement replaces the previous bilateral system of performance setting and monitoring and is the result of efforts to integrate the airspaces of member states. As mentioned, the ministry remains the major principal on behalf of the European Commission. The agreement has a complex provision that prescribes how financial shortages caused by underperformance on the part of LVNL have to be partly paid for by LVNL prior to being allowed to raise tariffs. Although this new agreement has gone into effect recently, it is too early to view the full impact this may have. In Box 6.1 some more information regarding the FABEC performance plans is provided.

Text Box 6.1   Performance management under FABEC (Source: Summary annual report LVNL 2011)

In 2010 the so-called Performance Scheme (EC no.691/2010) was adopted. This EU council regulation forms a major pillar of the Single European Sky: a system of performance management. From January first 2012 performance management applies to the en-route services. From 2015 performance management will also apply to the terminal services. At a national level or at the level of functional airspace blocks, the member states are to develop performance plans with objectives that are consistent with the EU-wide targets. In 2011, the FABEC states have jointly drawn up the provisional FABEC performance plan. This plan gives an insight into the joint objectives of the FABEC members on performance areas as safety, capacity and environment. The performance area of cost effectiveness is addressed in the national performance plans. The Dutch performance plan for the first reference period was submitted and assessed by the European Commission in 2011. It was decided that the plan meets the criteria set. The most important quantitative objectives for the years 2012- 2014 are:

• Reduction of the average ATFM delay per en-route and terminal flight to 1 mn at most.
• A capacity declaration at reliability of the first inbound peak (68 movements/hour at Schiphol Airport) of at least 92.5%
• Performance within the current standards of rules for enroute and runway operations.
• Reduction of the fixed single rate for en-route air navigation services by 3.3%.

The financial incentive tied directly to performance is a new element in the relationship with the ministry. In order to enable LVNL to deal with the newly introduced financial risk, it was recently allowed to keep its own equalization reserve. The new performance framework has only partly gone into effect yet and suffers from some technical challenges regarding attribution of results and controllability of the indicators. These problems were also encountered earlier in the bilateral arrangements with the ministry. Nonetheless, some positive impact of the FABEC agreement is reported from both LVNL and the Ministry. The mere existence of a financial incentive and the possibility for LVNL to keep an equalization reserve does give the performance dialogue a more serious and less ritual character. It is important to realize that this is reported for the dialogue between the ministry and LVNL and that for an air traffic controller these agreements, like the previous ones, do not have much influence if any on their daily work.

Actual use of PI in relationship ministry-LVNL
Although the policy dialogue with the ministry is highly valued by LVNL managers, the performance reporting to the ministry is not seen by everyone as very valuable. When assessing the PI that is reported to the ministry, one of the LVNL respondents noticed that:

'The introduction of key performance indicators for the ministry has not changed much since this is the information we traditionally used for our internal steering'

With regard to the FABEC performance targets and its financial incentive structure a respondent at LVNL noted:

'This will not affect our behavior in any way because we cannot influence the indicators, at least not on the short term. Maybe on a 8-10 year horizon one can.'

The ministry perceives the information that is reported about the FABEC targets as well as information on the noise level (through the Alders Conference) as very valuable to monitor progress and compliance as well as for its political process. In this sense performance reporting really does seem to fill (or at least ease) an information gap in technical expertise between principal and agent. The relevance of most PI for the actual budgeting process on the part of the ministry on a first glance seems to be rather limited. This results from the fact that virtually no public funds are transferred directly to LVNL. If one takes a second look however, the ministry, working closely with other stakeholders, has made a successful effort to curb spending of LVNL in support of the policy goals of the government. In response to the looming financial crisis, LVNL's projected incomes had to be adjusted downwards dramatically during 2009. Out of concern for the aviation sector in the Netherlands and the position of Schiphol airport, the Minister decided to restrict the tariff rate development for a number of years. This effectively prohibited LVNL from translating operating deficits resulting from lower operating volumes of air traffic, into higher tariff levels. To meet the reduced budget, LVNL set in motion an unprecedented cost cutting operation that resulted in the reduction of 128 FTE, over 12% of staff levels in 2008. The reduction in personnel did not affect the number of air traffic controllers as it was filled in entirely by the other LVNL units. Obviously the ministry did not bring about this efficiency operation at LVNL all by itself. Instead she was helped by other powerful stakeholders. On the other hand the ministry points at other member states that saw the costs of their air traffic service provider gone up in spite of the crisis.

It is interesting to determine the role of PI in this specific case. In terms of performance indicators, the minister decided to freeze LVNL's terminal unit rate at 2009 levels to control Schiphol airport's international ranking in in terms of aeronautical costs while facing a decrease in expected traffic volume figures (the number of flights). The traffic volume figures and Schiphol Airport's competitive ranking have thus been proven relevant PI in terms of the budgetary dialogue between LVNL and the ministry. One can argue that these figures are not that informative about the real performance of LVNL as an organization and that the ministry was just teaming up with other stakeholder to control cost prices, which happened to be in their joint interest. Indeed, some of the other indicators from LVNL's budget documents appear to be more relevant in assessing LVNL's performance but bear no

direct relationship to the size of LVNL's budget. Examples of these are: availability of capacity and delays on LVNL controlled flights. These indicators do gain more budgetary relevance in the new FABEC performance budget regime. Maybe atypical as an example of using PI to improve effectiveness and efficiency, the ministry, unlike many of its foreign colleagues, did choose to use its power as a principal to pursue its interest of controlling cost prices. While doing so it was well informed about LVNL's performance and seemingly successfully negotiated to make sure that performance levels would not suffer from the resulting budget cuts.

6.2.2   Performance management within LVNL

Internally, LVNL has performance measurement and reporting systems that belong to the most advanced in the public sector of the Netherlands. The most important systems include a framework that links LVNLs diverse outcomes such as safety, capacity and noise levels to its business processes. Operationally, the safety loop regarding incident registration and analysis is an important source for monitoring and improvement. The cascade of targets and indicators, both for internal management and for external reporting must be striking for anyone coming from outside the organization. This chapter will not offer a full description of LVNL performance measurement and management system. Nonetheless the most important elements will be described here briefly.

The basis for monitoring daily operations happens on work floor level by air traffic controllers who electronically report incidents and irregularities on a daily basis. The reports may contain anything from a car crossing a runway (runway incursion), airplanes entering blocks of airspace without prior permission (airspace infringements), aircraft flying too close to each other (breaking separation norms) to just a control tower being out of coffee. Other air traffic controllers analyze these daily reports looking for serious incidents. The serious incidents are further analyzed to find out what exactly happened, how it could happen and why it happened. The results of this analysis form the input for a constant process of improving and refining the operating procedures for air traffic controllers. In addition to incidents, there is daily reporting of other figures for amongst others flight activity and delays. These are consolidated and projected against the targets set in annual and business plans.

The incident reporting together with data logs from LVNL's technical systems (like for instance radar tracks) form input for the VEM performance standard, which is regarded the most important source of management and stakeholder information. The VEM acronym stands for Safety (Veiligheid), Efficiency (referring to capacity use) and Environment (Milieu) as it specifies LVNL's performance in these terms. In the VEM framework, the factors traffic demand, availability of airport infrastructure and weather conditions are taken into account as well as the interaction between them. LVNL uses the indicators and norms of its VEM framework to decide whether operations need to be adjusted as well as to inform

stakeholders about what they can expect from LVNL. On a monthly basis VEM reports are discussed in the management team. The framework originated at the turn of the century, when LVNL experienced that a useful dialogue with its critical stakeholders was hindered by the inaccessibility and technical complexity of LVNL's line of work. Today it is regarded by both LVNL and the Ministry as a shared framework that integrates the interests and expectations of both parties in an objective matter. Cost efficiency is not integrated in VEM as but this is seen as the next refinement of the current framework.

Of the different uses of PI mentioned in CH 2.3, almost all uses were said to occur according to respondents. The most frequently mentioned were:
- Performance reporting for external accountability
- Analyzing productivity and funding levels
- Identifying service problems and changing work processes

These were followed by:
- Setting program priorities
- Strategically reallocate internal resources
- Allocating internal funds

When talking to organization members, it becomes evident that PI is widely used at LVNL to learn and improve processes. This was illustrated by this example from one of the respondents:

'Once every quarter the performance indicators of one process are discussed in detail in the Management Team. When looking at incident statistics over 2010 two types of serious incidents turned out to occur most frequently. One of these was the category airspace infringements, which is one that I can't control by myself. After discussing with the Ministry an improvement task force with all relevant stakeholders was started to tackle this problem.'

'The other one was the category ground incidents. During our spring and fall courses for traffic controllers, we analysed each of these incidents to see precisely whether we made a mistake or another party did. Over 2011 this type of incident was no longer at the top of our list.'

The financial impact of using PI remains less clear except maybe for capacity planning purposes. The evidence also suggests that PI plays a role when assessing investment decisions. An illustration of this was given by one of the respondents:

'Our investment plan is part of the substantiation of the tariff levels to our customers. For example say that to KLM, each minute of delay is worth € 50. By multiplying this with the

number of flights we can calculate that an investment of € 3 million is expected to be worth € 20 million to KLM in delay of reductions. In that case the investment will be justified. Sometimes it is not. The initiative to replace the 1950s era Instrument Landing System by a modern Microwave Landing System. This did not prove to be feasible as it would cost KLM more in expensive modifications to their aircraft than it would save. That is why this initiative was abandoned after a short test phase.'

Not all respondents agreed that each investment decision is subject to a systematic cost-benefit analyses as this example suggests. In particular investments that benefit safety improvements seem to be analysed in this manner.

While reporting information for a learning loop system is considered a normal part of the job for air traffic controllers, this does not mean that at the indicators that LVNL uses are necessarily perceived as valuable to their daily work at the operational level. It is the expertise of air traffic controllers to apply the procedures in their daily operations. This is a complex activity where norms are sometimes conflicting and must be applied in a flexible way for the best overall results. One respondent put it this way:

'We are in a 24hr process and I plan 2 days ahead at a maximum. We are heavily dependent upon the weather when we are planning capacity at a given level of air traffic …'

'…Good performance indicators for holding air traffic controllers accountable for their work are impossible to find.'

Holding individual traffic controllers accountable for performance indeed is not part of LVNL's performance management system. Moreover the performance management is primarily aimed at continuous improvement of working procedures for air traffic controllers and at detecting and correcting possible performance shortfalls as early as possible.

6.2.3 Conclusion on PB implementation

At LVNL, an abundance of PI is used internally in a systematic manner to improve the organization's effectiveness (notably safety) and, to a lesser extent, for managing financial efficiency. However, performance is not coupled in a rigid or systematic manner to the budget. As mentioned output estimates like the number of flights do play a role in capacity planning and to some extent PI is used in business cases regarding investments. Although performance planning does seem to play a role in capacity planning and investment decisions, the bulk of LVNL's costs are determined at collective wage negotiations that are heavily influenced by LVNL's international peers.

The relationship between LVNL and the ministry has known a system of target setting and performance reporting for some time. The ministry, as LVNL's principal, uses PI reported by

LVNL to reduce information asymmetry and to actively monitor progress towards agreed performance targets. Although PI is not used in a direct manner to adjust LVNL's budget upwards or downwards, the ministry has used its power to curb LVNL's tariffs. For LVNL this meant that it was held accountable for the same performance levels in spite of lower incomes due to reduced air traffic. The dialogue preceding European target setting that resulted from FABEC agreements has ensured further professionalization of the performance dialogue between LVNL and the Ministry. Despite the provision of direct ties between performance and financial risks, under FABEC, it is too early to tell what the exact consequences will be for the performance dialogue with the ministry and the European Commission.

The formal system of internal performance management and performance reporting in the budgetary cycle is well advanced and arguably has no peer within in the public sector in the Netherlands. Moreover, despite the dazzling amounts of measured data LVNL produces and processes, all of it seems to have a purposes and seems to get used to learn and improve There is however some difference between the *de jure* linking of performance and budgets and its *de facto* relevance for budgeting. The perceived value of certain performance indicators may differ according to one's role within the organization. For the individual air traffic controller many aggregated data and performance indicators LVNL uses in its dialogue with stakeholders may not appear useful for their daily operations.

Table 6.1   Presence of indicators of performance budgeting implementation

Indicators		*Presence in LVNL case*
PB 1.1	PB is used by the agency in addition to traditional budgeting	Present to some extent
PB 1.2	PB is used by the principal to control the agency	Clearly present
PB 2.1	A high degree of de jure PB implementation	Clearly present
PB 2.2	A high degree of de facto PB implementation	Clearly present

According to the micro model of PB introduced in Chapter 2, these uses of PI count as successful PB implementation if they contribute to the phenomena of enlightenment and program learning, leading to more efficient and effective ways policy execution. The examples reported from LVNL respondents do seem to fit this criterion. The continuous improvement of LVNL's processes using PI should be regarded as program learning that is institutionalized to a high degree. Jasper Daams, in his 2011 study managing deadlocks in the

Netherlands aviation sector, offers a detailed description of several examples of double-loop learning by LVNL, or enlightenment in the model used for this research. One example is the handling of parallel departures after the introduction of a new runway at Schiphol airport in 2003. This cases describes the sequence of events in the 1996-2009 of LVNL and others stakeholders struggling to meeting the dual objectives of capacity enlargement for Schiphol and diminishing environmental impact while guaranteeing airline safety at the same time. Key to this case were the detailed quantitative standards for environmental effects and the shared responsibility for them that was accepted by the joint aviation stakeholders. In his careful analysis Daams classifies 6 of the recorded events as cases in which double loop learning by LVNL and its stakeholders took place. (Daams 2011: 156-186)

6.3 Exploring Neo-Institutionalist Explanations

The evidence does show that LVNL has been using performance measurement to enhance efficiency and effectiveness in performing its public tasks. However that does not automatically mean that this has occurred thanks to implementation of a PB system. To answer this question the results of the alternative explanations from the model have to be taken into account.

6.3.1 Explanations from Historical Neo Institutionalism

When investigating path dependency as an explanatory factor, important events with an impact the organization were sought that might explain PB implementation. This could be anything, for example a crisis with a principal or a series of bad press in the media. Moreover it is possible that, if adoption of PB is viewed as an appropriate response to such events, result oriented behavior should be explained by the logic of appropriateness rather than PB's incentives doing their work. To put it differently: organization members are susceptible to unwritten rules and roles that demand compliance with a performance management system, regardless of the incentives of this system itself.

To determine the occurrence of possible critical junctures, respondents in the interviews were asked to name the top 3 events that affected their organization throughout the years. Just like in the previous case study, they were asked to think of answers that they felt their colleagues might give as well. The latter remark was added to prevent that respondents mentioned events too close to their own specific duties. When analyzing the answers that that were given, a number of events become apparent have had some viable impact on the organization including its attitude regarding performance reporting and measurement. A total of thirteen events of a diverse nature were mentioned by respondents. Two of these were mentioned by five of the nine respondents:

1) The crash of Turkish Airlines Flight 1951 in 2009 at Schiphol airport while landing. The accident claimed nine lives. The investigation of the crash concluded that mechanical failure of the altimeter and the pilots' inadequate response caused this incident.

116

2) <u>The cost cutting operation that took effect in 2009.</u> This was the first time that the organization was faced with lay-offs effectively eliminating the notion of life-long employment at LVNL that many employees had

Another two events were mentioned by three respondents:

3) <u>The crash of El Al Flight 1862 on Amsterdam in 1992</u> while on approach to Schiphol for an emergency landing following mechanical failure. The Cargo Boeing 747 crashed into apartment buildings causing a massive outburst of flames. At least 43 people lost their lives including the 4 people aboard.
4) <u>An incident with a Delta flight aborting its take-off roll in 1998</u> when the pilots observed a towed Boeing 747 crossing the runway in front of them. This incident lead to legal prosecution of the air controllers on duty. All three were found guilty but a sentence was rejected after appeal to a higher court.

At first glance, the large number of diverse events mentioned, may not offer strong evidence of a large impact of a single traumatic event. When taking a deeper look however, that first assessment may require some adjustment. Firstly when applying a weighed score, corrected for the order of importance as ranked by the respondents, both crashes (events nr. 1 and 3) clearly do stand out. The reported impact of these crashes, although 17 years apart, was quite similar. In addition to heavily affecting professional pride of air traffic controllers, the aftermath of these disasters was characterized by critical appraisal of the performance and responsibility of air traffic control (LVB and LVNL respectively) in press, parliament and public opinion. The El Al crash even resulted in LVB employees being suspended from their job. Although the investigations later on demonstrated that these allegations proved false, these events, according to respondents, did highlight the necessity for LVNL become more open and transparent and inform and involve stakeholders despite the fact that its job is difficult to communicate to outsiders.

The impact of the austerity operation (event nr. 2), although possibly traumatic at the level of individual employees, does not seem uncommon for public organizations that are faced with massive cuts for the first time. An interesting aspect in the case of LVNL is that it will probably result in cost efficiency to be integrated into the VEM performance framework as a fourth element.

The Delta incident (incident nr.4) is reported to have had far reaching consequences on the performance culture at LVNL. The fact that air traffic controllers could be legally prosecuted when doing a bad job had a severe impact. On the one hand defensive routines became visible by a dip in incident reporting. On the other hand it contributed to the awareness that one needs to feel safe in order to learn from mistakes. With this in mind several measures have been taken to ensure a ´just culture´ to report incidents or failures and preventing

unnecessary exposure of employees involved in reporting incidents (see CH 6.3.2). This incident is indeed reported to have played a major role in the process of developing a 'just culture' as the aviation sector in the Netherlands perceived the prosecution as the introduction of a stricter policy and the example of the Delta incident was brought forward in many discussions later on (Daams 2011: 229).

When combining the consequences of these events for the organization and its attitude toward performance measurement and reporting, it becomes clear that incidents with airliners or near misses that got extensive media coverage dramatically emphasize the vulnerability of the individual air traffic controller. The response to this highlighted vulnerability has been twofold. Internally the events contributed to consciously protecting the vital process of incident reporting by meeting conditions like confidentiality and anonymity. Externally these events contributed to LVNL seeking a more proactive, fact based dialogue with stakeholders instead of being forced to respond from a defensive position. Most respondents agree that this strategy has been showing results as phrased by one respondent:

> 'LVNL used to get criticized more frequently because we were not dealing effectively with our environment. We can grow further by finding the optimal balance between self-criticism and maintaining the safe environment that is required for learning. This safety is also required from external parties'

Another alternative explanation for adopting a PB system might be the dominance of a certain influential unit or person that acted as a champion of output measurement or performance management. When looking at dominant parties within LVNL it hard not to notice that the craft of air traffic controllers is highly regarded within the organization and that (former) air traffic controllers in that sense form some sort of elite within LVNL. The VEM system that has been in use for about a decade now, was developed by the current CEO of LVNL, Mr Paul Riemens. Back in the 1990s when he was working as an air traffic controller he began modelling air traffic control for a thesis. Later on he was responsible for initiating this system and grew to become LVNL's current CEO. In 2011 Mr Riemens was appointed chairman of CANSO, the international board of ATC organizations. The fact that LVNL's primary performance management system was developed by the operational unit (and not by staff) as well as the career of the current CEO seem to offer a likely explanation for taking performance measurement seriously by organization members. One can argue that a system like this would only have made a serious chance of implementation if initiated from operations.

Finally, a dominant role of specialists in a certain policy field may explain why quantitative measurement is regarded higher than other, more political, factors. To determine this,

respondents were asked to choose between factors that they regard as most influential to LVNL policy in two subsequent questions:

A) Stakeholder interests   or   B) Political ideas and policy goals
and
A) Politics/ public opinion   or   B) Specialist norms

If for both questions answer B is chosen, this would indicate that LVNL's policy area is run pretty much by experts and their policy ideas. However, no clear picture emerged when applying this framework to air traffic control. The distinction between interests and ideals as the basis for public policy turned out to be a difficult one to make as policy interests of stakeholders are more or less translated directly into the policies conducted by the ministry. It is however clear that, compared to for example SBB, the environment in which LVNL operates is traditionally more dynamic and politically sensitive. This has to do with the large financial interests of a number of major players that have to cooperate quite closely with LVNL (e.g. KLM, Schiphol). Instead of designing policies for an agency to execute, the ministry acts more like a referee continuously balancing conflicting interests of different parties. Although the processes of LVNL are to a large extent quite technical and inaccessible to outsiders, this does not withhold other parties from having opinions on the outputs and outcomes of LVNL. For this reason a dominant role of experts in LVNLs policy area may offers a likely explanation for utilizing performance information. This does not mean that development of performance targets and measures takes place relatively undisturbed from stakeholder attention.

The overall presence of an internal champion of performance measurement in a powerful position does offer a likely explanation for the utilization of performance in formation as an agency. To a lesser extent this can be the case with the other historical neo institutional variables.

Table 6.2   Presence of indicators historical neo-institutionalism

Indicators		Presence in LVNL case
H 1.1	Critical juncture in the accountability chain or policy field	Clearly present
H 1.2	Problem to which PB was seen as a solution	Present to some extent
H 2.1	An advocate or champion of PB in a powerful position	Clearly present
H.2.2	A policy field in which specialists are dominant.	Present to some extent

6.3.2 Explanations from Sociological Neo Institutionalism

It can be argued that LVNL displays an almost compulsive commitment to improving airline safety. To a certain extent this commitment can be attributed to the entire airline industry. The results of these continuous improvement efforts have been obvious to everyone since the birth of this industry in the form of a spectacular increase in airline safety. An open culture in which important and also painful matters can be discussed openly is obviously a key factor to the ability to keep LVNLs commitment to learning and improving. The conditions for organizational learning were tested by letting respondents react to 4 statements:

a. Within LVNL, important issues are being discussed openly and fairly
b. Within LVNL, existing opinions are regularly challenged and discussed
c. If LVNL is being confronted with a problem, a thorough problem analysis takes place prior to taking action
d. If things don't work out in LVNL's policies and execution, lessons are usually learned

According to the interviewees these characteristics do match the culture within LVNL. In addition, LVNL management is highly aware of the so called soft side of maintaining a corporate climate where learning from mistakes can take place and also takes measures to actively promote this. Examples of such measures are minimizing physical and technical barriers to incident reporting, safeguarding discreteness and anonymity and praising employees for reporting problems. Again, this is something that goes back a longtime in air traffic control at an operational level and also in the airline industry as a whole. A lesson learned early on from analyses of airline incidents has been that crashes can occur when a pilot or controller refrains from disputing a superior's inadequate assessment of a situation. One respondent reflects:

> 'Calling attention to someone's behavior or mistakes has been rooted in the air traffic control culture for about 20 years. In the old days, the persons with the big mouth often got his way. This culture-shift has been going on for a long time at Operations. The rest of the organization lags a little behind in this respect'.

Other respondents confirm that when it comes to perfection of the procedural and cultural learning mechanisms, operations is in the lead. Insofar there is any hesitation this concerns the dominant position of the air traffic controller's profession within LVNL and the LVNL's occasional hesitation to open up to outside stakeholders. Some claim that due to the complexity of the job, one does need a certain level of knowledge to be fully accepted as a negotiating partner to an air traffic controller. Some critical reflections from respondents on LVNL's openness are:

'A downside of the self-critical attitude is that LVNL does not always accept outsiders' criticism easily. LVNL is perfectly able to critically self-reflect but outsiders should not necessarily do the same.'

'There is some fear that openness to the outside world will be punished unfairly. I am convinced that LVNL is open but we also want to be conscientious. That is why LVNL sometimes may appear like a closed bastion to outsiders.'

'This goes in particular for the outside world but, to a lesser extent also internally, when other units criticize Operations.'

It should be noted that most respondents who point to these problems also claim they have improved over recent years.

In contrast to the SBB case study, no opportunity existed at LVNL to conduct a large scale survey to validate the answers given in the interviews. When looking at the SBB case, the results from the interviews had to be adjusted downward after including the questionnaire results. It is possible that the responses from the interviews may also not be representative of the entire organization because of an overrepresentation of management and an underrepresentation of staff and operations. Also social desirability can play a larger role in the interviews. For these reasons, additional independent sources were sought to complete the assessment of LVNL's participative and reflective openness and culture with regard to learning behavior. One of these was found in the independent ISO quality assurance audit, conducted by *Det Norske Veritas*. In their audits for LVNL's ISO 9001-certificate it is stated that:

'A strong focus exists on learning from experience, events and assessments. This can be found on several levels in the processes concerning Management, Change and Delivery. The Plan-Do-Check-Act mechanism is present.'

'Management was able to show how progress was measured and monitored from performance measurement.'

One of the other sources studied to gain additional objectivity were an employee satisfaction survey conducted bi-annually. Relevant statements for respondents from this survey were:
- *When I see an opportunity for improvement, I take the initiative to improve*
- *LVNL encourages taking initiative*
- *As colleagues we can call attention to each other's behavior*

The responses to these questions showed that LVNL employees in majority agreed with them, although LVNL did not score significantly higher than its benchmark organizations from the public and private sector.

Combining these sources, the findings from the interview responses may indeed need some adjustment downwards. However this does not change the image of a corporate culture in which a high degree of reflective and participatory openness is clearly present.

Another important factor that enables the use of a PB system for organizational learning is the absence of distorting cognitive frames. Distorting cognitive frames hamper a learning dialogue because the PI that is generated and reported is interpreted differently by different members of the organization. Ideally, the fit between the formal performance measures that the PB system provides on one hand and the convictions of organization members of what is considered successful on the other hand, is a complete one. This would mean that organization members hold a shared view of the meaning and usefulness of measured performance data. This was investigated by asking respondents about their attitude towards performance indicators, their knowledge of LVNL's performance measurement and whether they believed it was beneficial to their organization.

Given the abundance of performance data collected at LVNL, it is hard to imagine that each organization member is aware of each performance measure, let alone hold a shared view of their meaning. Indeed the individual air traffic controller's interference with LVNL's performance management system is limited to his or her contribution to incident reporting and periodical involvement process improvements. Based on the answers from the interviews however, the use of measured performance data is seen as very useful by almost everyone and seems to be quite undisputed. This goes in particular for the incident reporting statistics for air traffic controllers and the VEM framework for managers. Especially the performance reports produced by the Strategy and Performance unit are viewed as indispensable for LVNL's managers. This does not mean that there is no criticism of some of the performance indicators LVNL has to report on. As one respondent explains:

'Working with indicators is widely accepted because the usefulness of some data is clear to everyone. This goes in particular for safety. This is somewhat different for environmental performance measures. If a single plane is in violation of the norms, it is impossible to attain your target for the entire year. This does not motivate air traffic controllers in their daily work.'

FABEC performance goals are also criticized by some for the inability of LVNL to influence certain outcomes as well as the opportunity that exists for moral hazard on the part of other air navigation service providers.

Concluding, the key elements of LVNL's performance management system are well known and viewed as relevant by organization members. From a viewpoint of sociological neo institutionalism, the ability to openly discuss problems and reflect on existing positions make for an organizational culture in which using of PI for learning and enlightenment is regarded as appropriate behavior. In other words all the investigated indicators can offer a likely explanation for utilization of performance information.

Table 6.3   Presence of indicators sociological neo institutionalism

Indicators		*Presence in LVNL case*
S 1.1	Absence learning disabilities	Clearly present
S 1.2	Participative openness	Clearly present
S 1.3	Reflective openness	Clearly present
S 2.1	Shared view of the meaning of measured performance	Clearly present

6.4   Contextual factors

The international community of the Air Traffic Management industry can be characterized as relatively cohesive and puts a strong emphasis on quality management and benchmarking. An example is the CANSO Fitness Check for Air Navigation Service Providers. This self-assessment instrument is based on ISO 9004:2009 Quality Management Guidance Standards and measures progress towards the desired future situation of member ANSP's on vital performance areas. Amongst the 19 criteria measured are: strength of safety culture, monitoring and measuring, improvement priorities, and improving competences. LVNL is one of 30 member ASNP's that use this instrument.

The close interconnectedness with the competitive international market of civil aviation may also be a relevant contextual factor for this case. The financial crisis and the low margins of airline companies provide LVNL with strong external pressures. It can even be argued that the primary incentive for efficiency at LVNL is in fact the external pressure of the main carrier of the Netherlands (KLM) to keep tariffs internationally competitive. It is interesting to note that goal alignment between LVNL's stakeholders (Ministry, KLM and Schiphol) brought it to adopt an extensive austerity program in 2008.

6.5   Conclusion

The investigation of the indicators for LVNL should provide answers to three questions to be answered for each case (see CH 4):

A. *To what extent is this a PB success?*

B *Are the identified conditions present?*

C. *Does context offer a likely explanation?*

The answer to the first two questions can be given when looking at the numerical scores, obtained from the interviews are expressed in the range of 0-2 and applied to the different indicators:
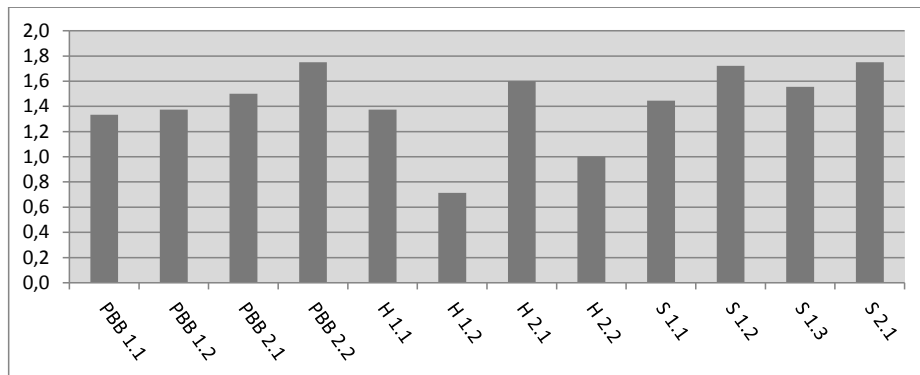
Figure 6.2   Numerical scores of the indicators from the interviews

The indicators for PB show that LVNL does fit all the criteria for a PB success. This is most evident for the *de facto* use of PI and, to a somewhat lesser extent, for the use of PI for budgetary purposes.

The second question should clearly answered positively for the presence of an internal advocate or champion of PB in a powerful position as well as for organizational learning, cultural openness and a shared view of the meaning of performance information. Less apparent but still present to some extent are the impact of events and the dominant role of experts in LVNLs policy area. Through the occurrence of sparse airline incidents in the Netherlands, the organization has been reminded of the vulnerability of the air traffic controller. It has responded by organizing more confidentiality in incident reporting while engaging in a more proactive dialogue with its stakeholders. The technical nature of LVNL work means that its recruitment results in a staff with a dominant 'engineers mentality' and a high tolerance for - and understanding of- quantitative measurement.

Contextual factors that may explain the use of PI by LVNL are the airline industry's commitment to safety improvement. This makes LVNL a major participant in an industry wide, data driven quest for improving its outputs and outcomes. It can be argued that the neo-institutional and contextual factors are likely to have interacted to a large degree here.

124

For example LVNL's preoccupation with 'just culture' is shared with other players in the aviation industry.

Looking at LVNL from a neo-institutionalist point of view, it can be characterized as an organization in which measurement performance to learn and improve and reporting performance to stakeholders are both regarded as appropriate behavior. This can partly be explained by traditions in the industry in which LVNL operates, its leadership and LVNL's organization culture. There may however be additional explanations that deserve some attention.

Epilogue
Like the SBB case, the intrinsic motivation of the air traffic controller is an apparent characteristic that surfaces from this case-study. Not unlike the forester, the air traffic controller does his or her work in a great degree of autonomy and has to rely largely on self-motivation. Further strengthening the professional autonomy of an air traffic controller is his or her extensive in-house education and a comparatively direct and dramatic impact of failure on the job. The autonomy aspects was illustrated by this quote form one of the respondents:

'Air traffic controllers are physically locked up during their work. They come in for their shift and immediately go home afterwards. That's why they operate rather independently from the organization's headquarters.'

The dominance of the obstinate independent professionals poses both advantages and disadvantages to the organization. On one hand he or she can be trusted to perform the job relatively unmanaged and be held accountable for it. In addition the expert is often passionate about the job and is motivated to share to professional knowledge to stakeholders and management. On the other hand, professional autonomy can be at odds with efforts of management or an organization's principal to control work processes or outputs, leading to occasional tension and claims of over controlling. Representative of this tension was this quote by one of the respondents:

'The aviation inspection is concerned about 10 general aviation incidents compared with about 400.000 departing commercial jets annually. Our job consists of 99% serious business and about 1% fun. If you take away that 1% of fun it will eventually have repercussions for the other 99%.'

Besides the autonomy and job clarity of air traffic controllers, the profession also has a long standing tradition of performance reporting as was shared by one respondent:

'Air traffic controllers have always known a self-learning norm. Nowadays we have a basic safety loop program, incident investigations and surveys to seek out opportunities improvement. In the old days there used to be a little hand written book in which the daily events were written down by the air controller on duty.'

With regard to their main performance management systems, a striking similarity with the SBB case is the fact that these systems started out relatively long ago and were initiated within the primary process. Gradually the systems developed into being the tools for central planning and reporting by management and staff. Subsequently the performance management systems gained an additional role for external accountability to a principal and other stakeholders. This pattern does hardly fit the archetype top down use of PB systems for goal alignment by a principal. Indeed, the control of government as a principal over LVNL with regard to the nature of its job should not be overrated as one respondent tellingly expressed:

'Government has not urged us to be occupied with airline safety, it is rooted in our organization and in our job. We are able to answer questions, the ministry is not even able to ask.'

Having read all of this it may appear as if institutionalizing the performance dialogue with the principal, performance based financial incentives and internal performance management have not have a significant impact on LVNL. The long standing tradition of performance monitoring and learning seem to come with the job of air traffic controller. Derived from this tradition are LVNL's PB systems that, unsurprisingly, fit the organization well because they were initiated from LVNL's primary process itself?

Depicting the PB system of LVNL as an entirely autonomous development by the agent himself may however not tell the entire story. For example the ministry does claim to use the information reported by LVNL to reduce information asymmetry and has successfully imposed spending cuts on the organization without sacrificing performance goals. Meanwhile LVNL's internal performance management has been developed into a framework that not only LVNL's management, but also its stakeholders view as indispensable.

So in spite of the presence of several favorable conditions explaining this PB success, formalizing the performance dialogue and performance management system may have added value for LVNL and its stakeholder. Both financially as in terms of the stakeholder dialogue.

Data collection LVNL case
The primary method of data collection were semi structured interviews. Between March and November 2012, ten persons were interviewed. Two of these worked at the Ministry of

Infrastructure and were charged with oversight on LVNL. At LVNL HQ eight persons were interviewed representing the Executive board and all executive and staff units except Legal Affairs (see Appendix II for more details). Data were primarily processed by qualitative analysis although some quantitative analysis was used for comparing and aggregating results (see section 4.6). These scores were aggregated for each case allowing for a qualitative comparison on an ordinal scale. After finishing each draft case study, the report was reviewed by a key informant from the organization involved. By means of triangulation the findings from interviews were compared with questionnaire results and findings from analysis of documents sometimes nuancing the findings from the interviews. As explained in Section 6.3.2, no opportunity existed at LVNL to conduct a large scale survey to validate the answers given in the interviews. As an alternative, internal questionnaires conducted for LVNL's quality management were consulted on the scores on key variables.

Sources

Algemene Rekenkamer 2004, *Toezicht op de Luchtverkeersleiding Nederland* (Tweede Kamer, vergaderjaar 2004-2005, stuknr. 29 825)

Algemene Rekenkamer 2007, *Toezicht op de Luchtverkeersleiding Nederland – terugblik 2007* (Tweede Kamer, vergaderjaar 2006-2007, stuknr. 29 825)

www.canso.org *Fit for purpose - Fitness Check provides roadmap to seamless ANS*

Daams, Jasper 2011*, Managing Deadlocks in the Netherlands Aviation Sector*, Delft Eburon

LVNL, *Annual report 2009*

LVNL, *Annual report 2010*

LVNL 2012, *Jaarverslag 2011 - Verantwoording aan de staatssecretaris van Infrastructuur en Milieu*

LVNL 2012*, Factsheet 1 mei 2012*

LVNL 2012, *VEM Performance Standard Winter* 2012/2013

www.lvnl.nl

Ministerie van Verkeer en Waterstaat 2005, *Evaluatie Luchtverkeersleiding Nederland LVNL* (evaluatie vindt elke 5 jaar plaats uit hoofde van art.5.46 van de Wet luchtvaart)

Ter Avest J.A., Clerc F. and Groenendijk, J.M. 2005, *Naar een optimale toepassing van het VEM Raamwerk - Evaluatie Performance Luchtverkeersleiding Nederland (LVNL) van het Veiligheid, Efficiëntie en Milieu (VEM) Raamwerk.* Report by Twijnstra Gudde requested by Ministerie van Verkeer en Waterstaat

Volkskrant, Feb 5th 1999, *Verkeersleiders kijken tv met het zweet in hun handen*

Volkskrant, March 5th 1999, *Een elitaire club van specialisten*

http://wetten.overheid.nl/BWBR0032482/geldigheidsdatum_09-02-2013

CHAPTER 7   CASE STUDY UNITED STATES FOREST SERVICE

7.1   Description of the Agency and its Principal

History and tasks USFS

The Forest Service (FS) is a Federal agency that manages public lands in national forests and grasslands. The service was established in 1905 to provide quality water and timber. During these days, the influential first Chief of the Forest Service, Gifford Pinchot, summed up the purpose of the FS '*to provide the greatest amount of good for the greatest amount of people in the long run.*' Over a century later the initial purpose still echoes in the Forest Service's motto: '*Caring for the Land and Serving People'.* Throughout the years the public have been demanding the management of additional resources from national forests and grasslands, notably forage for livestock, wildlife and recreation. Today the area managed by the FS amounts to nearly 30% of US federally managed lands similar in size to the state of Texas. The FS is also the largest forestry research organization in the world providing technical and financial assistance to government and private sector parties at home and abroad. The main areas of activity can be summarized as:

- Protection and management of natural resources on National Forest System lands. This includes a wide variety of activities by forest rangers such as treatment and restoration of habitat and watersheds, managing wildlife and protecting ecosystems against evasive species, prevention and suppression of wildfires, maintaining roads, trails and recreational facilities.

- Research on all aspects of forestry, rangeland management, and forest resource utilization.

- Community assistance and cooperation with State and local governments, forest industries, and private landowners to help protect and manage non-Federal forest and associated range and watershed lands to improve conditions in rural areas.

- International assistance in support of U.S. environmental policy

It should be noted that, like many public sector agencies, Forest Service's activities have some degree of incompatibility. Examples are balancing the economic interests of local communities and logging companies versus sustainable use of forest resources or encouraging recreational use of forests and safeguarding natural habitats and preventing wildfires.

Size

The FS currently employs around 30.000 people at some 750 locations. Most numerous are those working for the National Forest System and in firefighting (each about one third of the workforce). In the summer months the workforce is augmented by a seasonal workforce of

temporary contractors and volunteers to deal with the annual peak in workload from recreation and wildfires. Quantitative assessments of the reliance on volunteers are hard to come by. In 2007, the number of hours worked by volunteers nationwide was estimated to be about 2.500.000 which would be around 1.500 full time workers (Absher 2007). The annual budget is about 5.5 billion US$ (enacted budget 2012 – see Figure 7.2 for a breakdown). Revenues amount to about 0,5 billion a year with timber, minerals and recreation being the largest contributors.

Organization
The FS operates under guidance of the U.S. Department of Agriculture's Undersecretary for Natural resources and Environment. The Forest Service's national headquarters are referred to as the Washington Office or WO. It is home of the of the politically appointed Forest Service's Chief as well as to the Chief Financial Officer and four deputy Chiefs for:

- *The National Forest System* – tasked with resource management on the public lands.
- *State and Private Forestry* – tasked with cooperation with parties involved in the management, protection and development of forests that are not federally owned like private landowners and local government and forest industries.
- *Research and Development* - provides the scientific and technical knowledge in the field of forestry. Research is conducted through a network of forest and range experiment stations and the Forest Products Laboratory.
- *Business Operations* – responsible for staff activities such as acquisition, human resources and information resources

A number of other organizational units charged with international and research activities reside under the WO. The bulk of the organizational activities however take place in regions, stations and areas located throughout the nation and generally referred to as 'the field'. The field is organized into three hierarchical levels:

The Region: The US is divided up into 9 regions that each have a regional forester in charge. The regional office (RO) staff coordinates activities between national forests, monitors activities on national forests to ensure quality operations, provides guidance for forest plans, and allocates budgets to the forests.

National Forest: There are 155 national forests and 20 grasslands.. The person in charge of a national forest is called the forest supervisor. Forest supervisors of the national forests within a region report to the regional forester. The headquarters of a national forest is called the supervisor's office. This level coordinates activities between districts, allocates the budget, and provides technical support to each district.

Ranger District: Each forest is composed of several ranger districts. The district rangers from the districts within a forest work for the forest supervisor. There are more than 600 ranger districts. That vary in geographical size (20,000 to 400,000 hectare) as well as number of staff (10-100) Many on-the-ground activities occur on the ranger districts, including trail

construction and maintenance, operation of campgrounds, and management of vegetation and wildlife habitat.

Vital to understanding the rather complex organizational and management structure of the FS is its dual character. Apart from the hierarchy based on regional structure (Chief - Regional Forester – Forest supervisor – District Ranger) another line of authority exists that follows the programs within a particular policy area. (Deputy Chief - Program managers WO – RO manager for Strategic Objective – Program managers at National Forests).

Policy field and tasks principal
As mentioned the FS functionally resides under the U.S. Department of Agriculture (USDA). The US Federal government has three other agencies that are primarily tasked with managing federal lands and natural resources. These are the National Park Service, the US Fish and Wildlife Service and the Bureau of Land Management. These three agencies all fall under the Department of the Interior.

The FS formally responsible for executing strategic goal number 2 from USDA's strategic plan 2010-2015:

Ensure our national forests and private working lands are conserved, restored, and made more resilient to climate change, while enhancing our water resources.

This strategic goal is translated into 4 objectives and 15 associated performance targets:
Objective 2.1   Restore and conserve the nation's forests, farms, ranches, and grasslands
Objective 2.2   Lead efforts to mitigate and adapt to climate change
Objective 2.3   Protect and enhance America's water resources
Objective 2.4   Reduce risk from catastrophic wildfire and restore fire to its appropriate place on the landscape

Progress on these objectives is measured by 15 performance measures with associated baselines and targets, 10 of which are reported on by the FS. The other five are not directly address the FS responsibility as they refer to non-federal land (2.1 and 2.3), carbon sequestration (2.2) or flood prevention (2.3). Prior to the current strategic plan, the FS was tasked with strategic goal number 6 from the 2005-2010 Strategic Plan which stated a more or less similar set of goals and objectives.

Although operating in a relatively stable policy area for the last two decades, the activities of the FS have been impacted by some major policy changes, notably in the 1970s and 1990s. First of all the arrival of federal environmental legislation in the early 1970s (National Environmental Policy Act, Clean Air Act, Clean Water Act) saw a shift in focus to sustainable use the lands. The adoption of further legislation such as the National Forest Management Act of 1976 and elaboration of biodiversity legislation in the 1990s lead to disappearance of the controversial practice of massive clear cutting of forests. One of the controversies that

received quite a lot of media attention in the 1990s was the fight between loggers and environmentalist protesters for conservation of Spotted Owl habitat in the Pacific Northwest (also referred to as the 'timber wars'). These events resulted in a significant drop in timber harvest volumes cut from National Forests as illustrated in Figure 7.1.

Figure 7.1 Decline in timber harvest (Source USFS, visualized by EcoWest, www.ecowest.org/2013/05/28/timber-harvest-falls-in-national-forests)
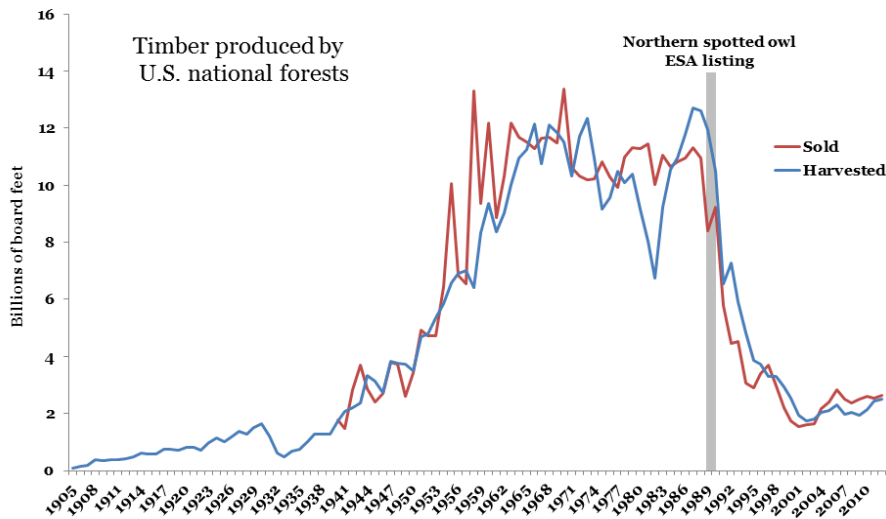
It can be argued that the field of Forestry in the entire western world and beyond has experienced a similar shift in focus from timber to sustainability and more diverse uses of land.

In response to these developments the character of the FS workforce also changed as specialists trained in other areas like for example wildlife biology started entering the workforce in larger numbers. In addition the FS has long had an image of being a service dominated by a white male culture that is consistent with stereotypes of lumberjacks and rangers. Promoting workforce diversity has therefore been a fairly consistent priority of top management throughout the previous decades.

Combining these developments the FS has been on a long journey to change from a white male dominated organization primarily directed at harvesting timber to a multidisciplinary land management organization with a diversified workforce. Although this development has been going on for a long time, reflections of the old versus the new culture were provided by several organization members.

Choice of principal and agent
Due to the variation in character and size of units within the Forest Service, for this case study a more specific focus had to be chosen. Principal-agency relationship can be defined at

multiple levels in the case of the FS. As described the FS is the executive agency for one of the USDA's primary objectives. In addition to the USDA acting as a principal towards the FS, both agencies are scrutinized by the White House's Office of Management and Budget (OMB) with regard to efficiency and effectiveness issues. Within the FS the WO manages relationships with all 9 regions to realize its goals. Within a region a number of National Forests report to the RO for accomplishing their tasks. A similar relationship exists between ranger districts and the supervisor's office of a National Forest.

Although the relationship between the FS WO and the USDA is taken into account in this study, most FS employees perceive 'Washington' as a single principal. Therefore the primary focus will be on the relationships within the National Forest System between the Forest Service's Washington Office and one of the nine regions. For assessing the internal performance management, the relationship between regional headquarters and a particular National Forest was selected.

For this purpose the regional Headquarters of Pacific Northwest Region 6 (R6) and the Gifford Pinchot National Forest (GPNF) were selected. Region 6 was selected because it had a good reputation with regard to performance budgeting according to the WO and was fairly representative for a region in other respects. The GPNF was selected because it is one of the major National Forests within Region 6 and for practical reasons because it was in close proximity to the RO Headquarters. In terms of size difference with the Netherlands Forestry case it is noticeable that the size of the GPNF alone equals more than twice the combined land managed by the Forestry Service of the Netherlands.

7.2   Degree of PB implementation

As the federal US never formally adopted a system of program budgeting, the Budget Line Items (BLI's) of the FS form the backbone of the financial allocation of the Service. These have been pretty stable for at least the last decade and are visible in the budget documents at all levels of the organization. The allocation between these BLI's has also been reasonably stable with the of exception Fire management which relative share continues to go up at the expense of Capital improvement and Maintenance (see Figure 7.2).

Figure 7.2   Relative allocation of funds to FS BLI's
the 2012 Wildland Fire Management BLI includes the FLAME Reserve Fund that was introduced in 2010 as a transfer account to prevent borrowing from other BLIs during a severe fire season.

The reliance on BLI's for allocation instead of lump sum financing around policy objectives has long been considered somewhat problematic from the viewpoint of budgetary flexibility at the level of districts (Kaufman 1960:123). Several initiatives have been developed to integrate different budgets to support single set of goals and accomplishments. So far none of these has led to permanent changes.

The budget process at the FS is characterized by a large degree of stability. Within the budgetary and performance planning process, both lines of authority mentioned (geographical and policy area) are involved as program and geographical managers jointly set targets and negotiate reallocation proposals. Several respondents who had been with the service for 30 years or more and who had served at different locations, made this point. Moreover, the description of the FS budget process in the late 1950s by Kaufman seems to be largely accurate of today's budgetary process. In fact Kaufman refers to the Service's output based budgeting system as *performance budgeting*, long before this term gained popularity in public sector budgeting (Kaufman 1960:112).

7.2.1   PB in relationship WO – the Field

The primary performance document for the Service the national level in recent years has been the FY 2007-2012 FS Strategic Plan. The Strategic Plan can be viewed as a detailed partial translation of one strategic objective on National Forests of the Service's principal, the US Department of Agriculture (USDA). In this plan the FS identified 7 strategic goals and their intended outcomes. The 7 strategic goals were each divided into 1 to 5 objectives with associated measured performance measures with baseline and target values.

To add to the complexity, there is another performance structure to report to with yet another time interval. These are the political priorities of both the President and the

Secretary of Agriculture. The current presidential initiative is *America's Great Outdoors* which seeks to encourage grassroots conservation initiatives by local communities in order to make the federal government a better partner with states, tribes, and local communities. In 2009 the Secretary of Agriculture formulated his '*All Lands*' vision for forest conservation. This vision stresses the role of healthy forests and grasslands in protecting water resources and increasing resilience to climate change. Although both initiatives certainly receive attention and resources by managers, they are less visible in the day to day set of performance measures and output funding on the district level. Largely in line with these national priorities, Forest Service's politically appointed chief brings her or his own emphasis. The present chief, not surprisingly, stresses water integrity and the *Great Outdoors* objective to involve more urban population more actively in the nation's forests. During the previous decade earlier Forest Services chiefs prioritized subjects like unauthorized vehicles, invasive species management, hazardous fuel and climate change.

Like other US agencies performance accountability has recently focused heavily on the extra funds allocated as part of the ARRA, the Obama economic stimulus program. The complex cascade of performance planning and goal alignment as of 2012 is illustrated in Figure 7.3.

Figure 7.3   Cascade of performance planning US Forest Service
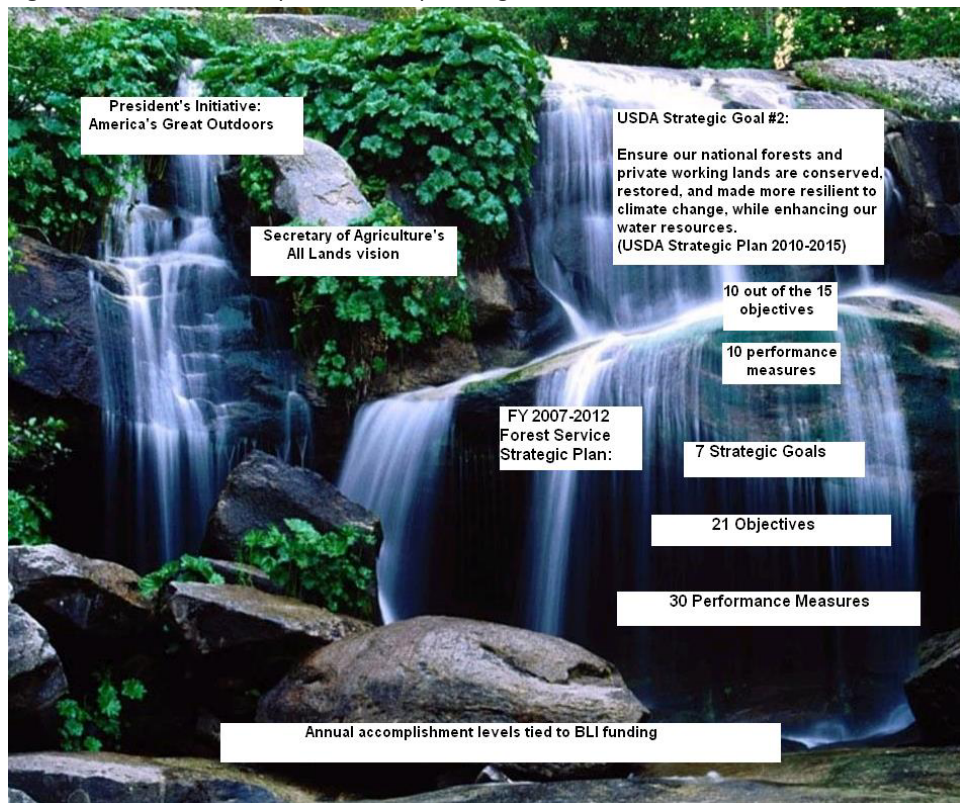
In recent Budget Justification documents, the Forest Service's budget was divided into the seven strategic goals from its strategic plan. Although both play a role in the performance planning and reporting structure, the BLI's seem to be the more dominant entities used for this purpose.

Figure 7.4   Division of FY 2012 budget into BLI's and Strategic goals (source: USFS FY2013 Budget justification)

2012 Budget by BLI	x 1 million
Forest & Rangeland Research	$295.30
State & Private Forestry	$252.90
National Forest System	$1,554.10
Capital Improvement & Maintenance	$394.10
Land Acquisition	$53.70
Other Appropriations	$5.90
Wildland Fire Management (incl. FLAME Wildlife Suppression Reserve)	$2,289.90
Mandatory appropriations	$738.70

Total
5.584,6
Million US$

x 1 million	2012 Budget by Strategic Goals
$4,020.90	Goal 1: Restore, Sustain and Enhance the Nation's Forests and Grasslands
$223.40	Goal 2: Provide and Sustain Benefits to the American People
$279.20	Goal 3: Conserve Open Space
$502.60	Goal 4: Sustain and Enhance Outdoor Recreation Opportunities
$390.90	Goal 5: Maintain Basic Management Capabilities of the FS
$55.90	Goal 6: Engage Urban America with FS Programs
$111.70	Goal 7: Provide Science-Based Applications and Tools for Sustainable Natural Resources Management

At the WO budget preparation starts with deputy chiefs setting directions for their areas of expertise with the deputy chief for the National FS typically taking the first shot at setting national and regional targets for the 9 regional foresters. Program directors at the WO and regional managers see if these targets are reasonable and proposals for adjustments go back and forth for a while. The agreed upon levels of funding and output are then put in the Forest Service's budget request which is part of the USDA's Budget request. The metrics tied to the line items are translated down to the National Forest level. It is not until the final budget resolution by Congress that the FS is sure about its definite spending levels and associated accomplishment levels. For most of the previous decade this wasn't settled until

halfway the fiscal year or even beyond. When definite figures and possible adjustments to the proposed budget are known, the WO passes down the adjusted budget figures and associated output levels (referred to as accomplishments in the FS). The RO gets assigned a regional target for each of the accomplishment measures it has to report on. These are subsequently allocated to the Forests and areas that reside under the region. Awaiting budget authorization, the regions and National Forests have to work with conservative estimates based on previous years.

Accomplishments are fed directly into databases from the project level using a variety of systems. The database can be accessed for monitoring progress against the targets by all levels of the Service and allows for a systematic drill down to regional and project level. Currently this system is being enhanced by introducing more visual geo spatial display possibilities. The vast majority of metrics used are on an output level although the number of outcome measures is slowly increasing (e.g. data on the condition of watersheds). From the point of view of the WO and the RO, foresters and forest supervisors are granted considerable freedom to negotiate target and funding reallocation as long as regions meet their targets. Each of the nine Regional Foresters provides an annual accomplishment report that is used in his or her yearly evaluation. This report is accompanied by a report in which the Regional Forester certifies that he or she has confidence in the data and that the controls used are valid. To safeguard data integrity an internal review process is in place at the Regional level. Under this system program managers in two forests each year have to prove that the numbers that were reported for a number of measures were actually accomplished. From the WO a more thorough review takes place every other year in a region. At all levels the general rule is that if a target is underperformed by more than 5% the manager will have to provide a written explanation has to be provided explaining why the target wasn't met.

Traditionally there has been a strong relationship between BLI's and accomplishments in the FS which was already described by Kaufman in the late 1950s and still captures the essence of today's budget formulation process (Kaufman 1960: 107, 115):

'The Forest Service budgeting is inextricably intertwined with a highly developed system of work measurement and planning….. There WO maintains a standardized list of all specific jobs – that is, discrete operations characterized by countable work output at specified standards of quality (perfection and intensity) – that must be done to perform all of the functions in national forest administration. It also keeps on hand records of 'converting factors' – the amount of time needed to perform one unit of work on each job. Finally, it has figures on the number of units of work on each job. Multiplying the volumes of business by unit time allowances yields the number of man-hours required to perform all the functions involved in managing the national forests. The man-hours are converted into dollar sums, and expenses for non-personal services (travel, supplies, communication etc.) are computed

as ratios of personal services. Thus, the estimates are formulated in terms of physical performance.'

Standardized units of output, cost calculations per unit and volumes are still the building blocks that determine how much of which output can be realized for a particular amount of funding.

The Forest Service's complex strategic performance planning seems to be largely separated from the BLI's centered budgetary process. The output planning tied to the BLI's can however, to some degree, be considered as a common denominator between strategy and performance, resulting in a clear performance budgeting and accountability system. This goes in particular for those objectives and performance measures from strategic planning have a pretty straightforward relationship with accomplishments levels as funded with a BLI's. This for example is the case with the target regarding timber sold (see Figure 7.5).

It should be stressed that the relationship between strategy, targets and dollars is stronger in some areas than other. For timber and hazardous fuels it is quite strong. For some targets it is not because BLI's can support a number of different accomplishments leading to complex cost attribution. A single accomplishment can also be funded by several BLI´s making it hard to specify the results of a specific type of expenditure The GPNF Work program for 2012 provides examples of both. Its recreation program is funded by nine budget line items and corresponds with different seven national accomplishment codes and associated targets. Only two of these national accomplishment codes which are linked directly to targets form both the FS Strategic Plan and the Budget Justification. From the remaining five, three are linked to targets from the Budget Justification while two aren't linked to targets from either document. Another example is provided by the GPNF watershed and Invasive species program which is funded by 3 BLI´s and corresponds with 7 accomplishment codes. 4 of these can be linked directly to the targets from both the FS strategic Plan and the FS Budget Justification. The other three to one of both documents only.

Looking at the entire system of accomplishment codes monitored by the WO, a similar pattern can be found. The WO monitors about 300 quantitative performance metrics, 121 of which show up in the work plans at unit or project level. As illustrated in Figure 7.6 only a minority can be traced to both a specific budget line item and a strategic objective.

138

Figure 7.5 Planning, monitoring and reporting system for amount of timber sold based on BLI funding levels (CCF stands for hundred cubic feet)

FY 2012 Forest Service Budget:

Budget Line Item:
National Forest Service - Forest products
$ 335.511.000
FTE: 2878

Associated Target for Forest products:
Volume of timber sold (CCF) 5,232,000

FY 2007-2012 Forest Service Strategic Plan:

Goal 2: Provide and Sustain Benefits to the American People
Objective 2.1: Provide a reliable supply of forest products over time that (1) is consistent with achieving desired conditions on NFS lands and (2) helps maintain or create processing capacity and infrastructure in local communities.
Performance Measure 2.1.1 : Amount of wood fiber provided each year to help meet the Nation's demand for forest products in an environmentally sustainable manner.
2006 Baseline: 5.4 million CCF
2012 Target: 8.0 million CCF, later reduced to 5,2 million CCF due to, amongst others, reduced demand because of the mortgage crisis

Input

Input

WASHINGTON DC OFFICE Source: Performance Accountability System

1 of 300 performance measures monitored: Vol. of timber sold (CCF), **Accomplishment code**:   TMBR-VOL-SLD

Annual Targets

Perfor mance dialogue

Data reported daily at project level

PACIFIC NOTHWEST REGION 6
Source: FY 2012 State of the Region report

Objective 2: Provide and Sustain Benefits to the American People
Objective 2.1.1. Forest Products

FY 2012 target volume 1,214,750 CCF
FY 2012 realization: 1,147,404 CCF

94% target attainment accomplished with 14% fewer dollars

Annual Targets

Perfor mance dialogue

GIFFORD PINCHOT NATIONAL FOREST
Source: FY 2012 Program of Work review

BLI Nation Forest System - Program Area Timber

Planned budget: $ 1,725,000   Actual budget: $ 1,740,658
Planned FTE: 19.34   Actual FTE: 17.94
Assigned target: 55,842 CCF   Accomplishment: 64,064 CCF

115% of Target completed, accomplished with 101% of funds utilized

Figure 7.6   FS wide performance measures and their relationship to the Budget and the Strategic Plan

As much as some might like to see a comprehensive structure where all activities have targets that are strategically aligned and have a unique cost attribution, the reality of the forest ranger in the field does simply not always allow for such comprehensiveness. The activities undertaken by a forester to restore stream habitats may also serve other goals such as protecting water resources, reforestation or countering invasive species.

For yet other goals there seems to be no good quantitative measure available. This is the case with goals like workforce diversity, recreation or community engagement following from the President's initiative. This is expressed by one respondent:

The targets are fair for what Congress appropriates dollars for. But there are also unfunded mandates, like are you an employer of choice or how we engage with communities. We just have to achieve those expectations.

Nonetheless, the Forest Service's immense and laborious system of codification, reporting and monitoring bears testimony to the fact that the Service attempts to make these links as explicit as possible.

7.2.2   Performance management within the US Forest Service

As described the WO can rely on a well-established system of measurement, reporting and analysis and, as a result, has a vast amount of information to its disposal to monitor regions and forests. The next question to address is to what extent this elaborate system is indeed used to improve efficiency and effectiveness of the FS in the way that was intended by performance budgeting reforms in the public sector. Of the different uses of PI mentioned in CH 2.3, all uses were said to occur according to respondents. Five of them were mentioned most frequently:

140

- Setting program priorities
- Strategically reallocate internal resources
- Allocating internal funds
- Performance reporting for external accountability
- Analyzing productivity and funding levels.

This was followed by
- Understand the impact of external events on performance goals
- Rewarding staff

Concerning the latter it should be added that there is no formal pay for performance scheme in place but an example from the field was mentioned were employees were rewarded cash amounts for over performing on already ambitious targets. For a further description of actual performance management a separation is made between the WO and the field.

Performance management by the WO
The regions regularly provide accomplishment reports to the WO but the most important one is annual one at the end of the year. Here the rule applies that underperformance in excess of 5% needs to be explained. Sometimes repeated failure to meet targets does lead to consequences like decreasing a target.

The Forest Service's external reporting consists of only a selection of the target levels that are managed by the WO and regions have to report on. There is indeed a distinction to be made between the targets that are reported externally and those that are used for internal steering, as one respondent reflects:

'We have 300 measures and only some 150 of those are shared externally. These are communicated in 4 places: in our budget justification, our annual performance report, USDA annual performance report and the USDA budget summary and performance plan. If you have a measure in those and you miss it there'll be analysis what is going on.'

'If you're not in there and you miss your target there is not much follow up. These other measures do serve a purpose for internal program management (have we done what we were hoping to do). ….We have been encouraging to limit these because it is a lot of work. There's been a notion for a while that your program only matters when there are measures.'

As Congress traditionally seems to be mostly interested in fire, timber and to a certain degree watershed, the other targets receive relatively little political attention. Fire obviously has a tremendous direct impact on local communities as does timber for its revenues partly go back to local communities to fund public services like schools. Watershed measures are important too but the complex causal relationship between for example water quality and

timber harvest is hard to communicate to the broader public. Special attention is paid to five key performance measures for USDA that are monitored and reported quarterly and also have quarterly targets. The quarterly reporting is a requirement from the recently implemented new federal performance management guidelines (GPRAMA 2010).

Apart from a performance dialogue between the WO and the field, USDA and OMB also monitor the targets the FS albeit with different perspective and attitude as a respondent explains:

'We do get a lot of tough questions about efficiency and effectiveness. Mostly from OMB and sometimes from USDA ….There are questions about performance form the USDA but they seem geared toward sharing information versus trying to change the way we go about our day-to-day business. We inform them if we take measures out of the measure set. They help us coordinate with OMB and OMB helps coordinate towards Congress. We try to keep them aware about what's going on the ground. USDA and FS keep each other posted to make sure that we do what the Secretary wants us to do. The USDA tries to get awareness in order to have a meaningful debate.'

The WO formally needs to approve every change in reallocation of targets but handles this with certain degree of pragmatism. Regional program and budget directors regularly make agreements and inform WO about this knowing that formal approval of reallocation will not be a big issue as long as national levels are attained. The WO's priority is that national levels are attained and an auditable record is kept about reallocations. A special provision exists for funding firefighting. If one of the regions spends more than they were allocated to fight a fire the difference is covered by the Flame Reserve fund that was established in 2010.

From the WO point of view, a consequence from the physical distance to the field is that local knowledge and data to analyze why a target is missed is often lacking. This leads to questions being asked to the field. One long serving respondent noted that the current advanced information systems did result in a reduced number of questions from Washington but did not reduce the total time spent reporting:

'Time spent on reporting is about the same as when I started 30 years ago. Back then if I wanted to see if a work was accomplished I literally had to go to a district's office and look at their file. Everybody believed what you reported because there was no really way of double checking. Now thanks to our big computers and our corporate database, someone at the WO can bring up a map of my forest and say 'show me where they did all the thinning for the last 20 years and it'll pop up and show him. As a result they don't call anymore with questions. 90% of the information request they can pull from the computer. For some specific information you have to go back to your files, for instance to analyze the impact of a certain decision.'

More generally the relationship between the WO and the field often appears to be a remote one. One respondent from the field noted that Washington was like a black box to them half of the time. An additional limitation for using the WO's databases for up to date monitoring is the timely entry of accomplishments. Throughout the FS the period from June until September is known as 'field season'. This means that the WO, the RO and Forest supervisors will start looking closely for lacking accomplishments to figure out whether they were not accomplished or they were not reported yet. Indeed it is acknowledged that completeness of data can sometimes be an issue. One respondent notes:

'Historically in the Forest Service, performance data is entered in the 3rd and 4th quarter and mostly in the 4th quarter and mostly in the last week of the 4ths quarter. It is a lot of work and people are busy, especially in the summer months when fire season is busy.'

To motivate foresters to timely entry their accomplishments, the rule was introduced that forests and regions are not credited for accomplishments that are not entered in time. In addition it is hoped that development of geo spatial reporting and presentation system will be a stimulus to timely enter accomplishments.

Performance management in the field
The budget formulation process as described so far may appear to have a rather top-down character. This is only part of the story however. During budget preparation phase in Region 6, the bottom up involvement in the process varies according to program manager as one respondent explains:

'We have 17 National Forests and 1 National Scenic Area. It varies by program and the program manager here at the RO. Some look at the target for the region and set targets based on what they think the region can accomplish. Some have conference calls, some have a more formal process where each forest would send their requests.'

Region 6 has a reputation of taking performance budgeting quite seriously. A reason for this is the development of a local performance budgeting process in which past results play a role in the distribution of next year's budget among the region's forests. According to this system, which was implemented in 2007, a unit can slowly increase its share of the total budget by repeatedly accomplishing more than their targets. Oppositely, continuous underperformance is punished by a decreasing share of the budget. The system is well balanced and can have a serious impact on forest management.

For allocation of the timber dollars line item within R6, a model is used with 3 criteria:
- Resource capability: the opportunities to do harvesting

- Timber infrastructure: mills, harvest companies, clients. This is important because some regions lost theirs and as a result there's no one to contract anymore.

- Past performance. This counts for 50%: a 4 year running average of real production against volume of financed target.

On the use of this system a respondent explains:

'After 4 years of falling of timber component in a particular forest, the director and staff went to this forest to discuss the program direction to see if the function of the forest should be shifted from timber harvest towards recreation. Since a new Forest Supervisor had been hired, it was decided to continue the timber program at the same level for another year and then re-evaluate once again.'

It is important to also note that not all metrics in R6 have this performance component and that the opinions on the effectiveness of this system differ as expressed by some respondents:

'At the start everyone was really excited about it. As our budget started going down (also from 2007) this makes it harder to actually see there has been an increase. In addition the increase is only small increase over time. If someone improves in this year it may take a couple of years to see a significant increase.'

'It works best in the timber area in my opinion. In some of the others like recreation they did not put in a performance element because they couldn't come up with the right metric. So recreation gets its funding based simply on the number of acres they have to manage.'

'In the last 5 years I exceeded targets in wildlife and invasive species by 50% and my budget is still going down because the overall budget cuts.'

Not surprisingly, the RO staff also has to deal with the technical performance measurability issues as they occur. It is up to the RO to explain the proper measurement methods to those doing the measurement following the guidelines provided by the WO. Two examples given by a respondent illustrate the kinds of issues at play here:

'We have stream habitat enhancement and lake habitat enhancement. One is measures in acres and the other in miles. If you improve land around a lake do you count the whole lake or a certain amount of feet around the area that is treated? We try to implement the best guidelines available.'

'How should landscape restoration be defined exactly? What do you count? If you add up all the acres you end up with more acres than we manage because it is a layered treatment.'

144

When it comes to monitoring at the regional level most units have program of work meetings where leadership will go over budget and targets. A month after the budget is published, forests have to put in their targets in the database. Sometimes it is more than the RO allocated and sometimes it is less .Depending on what that looks like the RO will monitor it. If it is much less, a conversation with the program manager may follow. There can be reallocation between forests during the year (so if one lags behind the other may take on extra work). In the early stages the monitoring is about what is planned. In the summer it is less about the planning but about reporting of the accomplishment.

In January the State of the Region report is published in which the progress for each strategic goal is described including successes, key issues and challenges for the coming year. According to a respondent the presentation of this report is a very successful annual event:

> *'It is a celebration of all the good work that got done. Forest supervisors look forward to this. Each objective gets 5 minutes and 2 pages. It is not the gotcha kind of evaluation. The only thing that does not work is that some try to use it as a way to get more money.'*

In addition to measurability issues and the administrative burden, another complicating factor for the performance budgeting system to work is the earlier mentioned fact that the field usually has to wait well into the fiscal year to know their definite allocation of budgets and targets. At the National Forest level the FS has learned to cope with budgetary uncertainty using an ingenious system that seems to have been in use throughout the regions for a long time (see Text Box 7.1)

At the GPNF program managers usually look at the budget data every month or quarter but are more frequently involved in accomplishment reporting using a number of different systems. It is not uncommon that one particular employee uses four separate electronic systems for accomplishment reporting. Program managers also review the data others put in the systems. There is a strong incentive to enter targets timely into the system as only planned workloads are funded when it is queried by the RO at the start of the fiscal year. All accomplishment reports have to be loaded into the database by Oct 30th or the forest will not credited with the accomplishments.

The incentive mechanism from the Region's performance budgeting system has also been felt in the past for timber at the GPNF as a respondent explains:

> *'There are financial penalties for the forest if we don't meet our targets. For example: several years ago we didn't meet our timber targets 2 years in a row. Because of that our forest allocation went down from 4.5% of the regional allocation for timber for timber to around 3%. So there was a 33% drop in fund allocation but the targets stayed the same. If you perform very well you get a bigger share next time around.'*

Text Box 7.1 Budget preparation procedure at Gifford Pinchot National Forest

National Forests have been facing a history of budgetary uncertainty with regard to delayed budget resolution and authorization by Congress, often halfway the fiscal year. At the FS the budgetary process at the local level has dealt with this uncertainty in a way consistent with the Services' long standing performance budgeting tradition.

- The overall budget baselines are calculated according to last year's budget level minus 5% 10% or more depending on the outlook

- First of all funding for necessary services such as permit administration and managing recreational facilities is estimated

- A second layer of funding and activities is defined explaining what could be accomplished with a little more money. Usually these are things that don't necessarily need attention each year like landscape or wildlife.

- Thirdly program managers explain what special projects could be accomplished if they would get some more money on top of that. Program managers are encouraged to come up with their priorities and plan at least 25% above the funding baseline. These projects are competitively presented to management who will decide on what projects to fund and prioritize.

The underlying idea is that if the Forest receives more than the conservative estimate, management does not have to go back and figure out what to do with it. If the Forest gets less funding, the line just moves up to the most urgent priorities. Essential in this process is that management oversees that the projects presented fit within the given priorities and that cost estimates are not too optimistic. The particular procedure above is the one currently used at the Gifford Pinchot National Forest and may vary according to place and time. Generally however is reflects a methodology that has been common at many National Forests.

Neither does the pressure from performance monitoring by the higher organizational levels go unnoticed at the Forest level. One respondent explains:

'Sometimes we have to talk to the region if accomplishments are lagging behind. In October I print out our target spread sheet including every target that is 5% or more behind. The supervisor has to write a report and may have to explain to the regional office.'

But here also the matter of selective target monitoring and large administrative reporting burden is mentioned by respondents:

'We have 30 or so targets that we have to hit but there are only handfuls that are real showstoppers (like timber or restoration). Additionally, funding might follow our accomplishment of things in those areas. Other targets may get lost in the mix because they don't have the necessary social pressure on them.'

7.2.3 Conclusion on PB implementation

The Forest Service's system of linking outputs to funding and labor has been stable over decades and was used well before performance budgeting came into fashion across the federal government. Its level of detail and design astonishes new employees to the service. Drill down functionalities will be further enhanced by making more use of digital presentation options. The accomplishments that are tied to funding levels form an avoidable part of the work of virtually all employees and form a stable foundation under each year's budgeting process.

Although the formal role of the WO in the Forest Service's budget preparation and results accountability seems to be strong in a formal sense, its actual performance management seems to concentrate on a selection of targets rather than overseeing comprehensive strategy implementation. More in general the WO seems to fulfill the role of the Service's gatekeeper towards its other DC principals and guardian of the Service's nationwide performance reporting system rather than micro-managing the regions and forests. It could be argued that the level of standardization of accessibility of PI combined with a disciplined workforce allows the Service to leave significant authority to the regions and forests for attaining their targets. For the internal management targets and data, the WO as well as the RO seems to rely strongly on the people in the field to make the right choices and will only intervene when national or regional targets are not met. The emphasis on decentralized authority has a long history in the Service as was noted by Kaufman half a century ago when citing a letter from 1905 by the Service's first chief Gifford Pinchot (Kaufman 1960: 84):

In the management of each reserve (now called National Forests) local questions will be decided upon local grounds…General principles… can be successfully applied only when the administration of each reserve is left largely in the hands of local officers, under the eye of thoroughly trained and competent supervisors.

When looking at the Forest Service's entire system of performance planning, reporting, monitoring and accountability, the goal alignment and monitoring intensity may look fragmented as it varies according to policy goal and field of activity. The WO does gather information on virtually all of the services activities and outputs in the field and has a solid accountability provision when targets are not met beyond a 5% margin. There is however reluctance towards intense monitoring and steering as this happens quite selectively. Indeed throughout the organization a rather clear distinguishing line exists between those activities that are reported externally and receive political attention (like fire, timber, watershed and USDA's key indicators) and the others. From the principal's perspective however the priorities seem to be covered quite comprehensively by the WO system. The overall image is one that the principal (the WO) uses its advanced PI and accountability system to control the agent (the field) regarding his priorities but is reluctant to use it for micro management of other outputs which are left to local managerial discretion.

Table 7.1   Presence of indicators PB implementation

Indicators		*Presence in USFS case*
PB 1.1	PB is used by the agency in addition to traditional budgeting	Clearly present
PB 1.2	PB is used by the principal to control the agency	Clearly present
PB 2.1	A high degree of de jure PB implementation	Clearly present
PB 2.2	A high degree of de facto PB implementation	Clearly present

The Forest Services' PB system can be criticized in the sense that not all of the services' goals seem fit well into its system. It seems like the method that worked well for timber has been adopted for all targets. For some this does not work out that well. For example a target for the completion of a land management plan is either reported as 0 or 100%. In addition the point can be made that more advanced performance measures linking outputs to outcomes seem to be underdeveloped and data integrity seems to be a reoccurring challenge. The Government Accountability Office (GAO) has been mentioning these points ever since the implementation of the Government and Results Act (GPRA) at the start of the century. Although this criticism is by no means pointless, it should be taken into account that given the level of sophistication and thoroughness of the Forest Service's PB system and the consistency and discipline it requires, there are limits of what can be expected. The reality of the foresters' work on the ground and the sometimes intangible wishes of political principals would provide formidable administrative and goal alignment issues for any good working PB system.

7.3   Exploring Neo-Institutionalist Explanations

As dominant as performance measurement, reporting and PI use may be in the Forest Service, it is not obvious that adopting a well designed and implemented PB system forms the main explanation for this behavior. As we know there are many examples where such systems fail to make a lasting impact on organizational behavior. In an attempt to understand the Service's preoccupation with performance budgeting and performance management the possible institutional explanations as introduced in Chapter 4 will be tested.

7.3.1   Explanations from Historical Neo Institutionalism

When investigating path dependency as an explanatory factor, important events with an impact the organization were sought that might explain PB implementation. This could be anything, for example a crisis with a principal or a series of bad press in the media. Moreover it is possible that, if adoption of PB is viewed as an appropriate response to such events, result oriented behavior should be explained by the logic of appropriateness rather than PB's incentives doing their work. To put it differently: organization members are susceptible to unwritten rules and roles that demand compliance with a performance management system, regardless of the incentives of this system itself. In this case events that affected the organizational attitude towards performance budgeting and performance management could be particularly revealing.

To assess the relevance of critical junctures in explaining the use of performance information, respondents in the interviews were asked to name the most significant events that have affected the FS throughout the years. Just like in the other case studies, respondents were asked to think of answers that they felt any colleague might also give in order to prevent answers too close to one's own specific duties. Subsequently the relevance of these events as possible explanation will be assessed.

The analysis of the answers that were given learns that no single large event clearly stands out as the two most frequently mentioned events were named by only 4 out of 8 of the employees that were interviewed:

1. <u>Political appointees from 1990s</u> – Under President Clinton, FS Chiefs (the head at the WO) became political appointees whereas traditionally they were promoted from the ranks of the Service itself. This has been feeding concerns of politicizing of the Service's top management.
2. <u>Adoption of ecological laws in 1990s</u> - This was preceded and followed by bitter confrontations between loggers and environmentalists also referred to as the timber wars or spotted owl controversy. This fell particularly hard on the FS in R6 as they had to navigate between maintaining national laws and their agreements with the timber industry and the local communities that dependent on timber revenues.

Two events were mentioned by 3 respondents:

3. <u>Centralization of Human Resources.</u> In 2006 the FS started to centralize administrative functions such as HR and IT into one shared service in Albuquerque, New Mexico. Apart from the impact on employees working in these functions and some initial start-up problems with service delivery, it also affected the sense of autonomy of units in the field.
4. <u>Environmental Legislation 1970s (clean air&water, NEPA).</u> This can be considered as the start of a process of more intense interference from Washington in the work of

foresters in the field including a rise in formal documentation and reporting requirements.

Another eight events were mentioned by just one or two people. One should bear in mind that the second event (Adoption of ecological laws in 1990s) is likely to have had a disproportionate impact on R6 as the Pacific Northwest was at the centre of the environmental protests at the time. Nonetheless the fact that the Service's policy execution was controversial and under public scrutiny is likely to have influenced the move towards increased public accountability as it has in the other cases. This is suggested by some respondents but may be considered a manifestation of a general trend more than an event specific to the Service.

When adjusting for the order in which respondents ranked the importance, events 1 and 3 stand out just slightly more. With regard to the first event (political appointees), extra emphasis of the priorities and initiatives of political leaders such as the Secretary or the President can be seen as one of the reasons for the Service's complex strategic planning as was demonstrated in Section 7.2.1. The uncertainty over this development was expressed eloquently by one of the respondents:

> 'The Chief used to be a civil servant who grew up in the FS. That is not true anymore. Now we get Chiefs and WO staff who never set foot on a National Forest. The FS grew up with the message that all our eyes should be on the National Forest. Now there are so many eyes in the WO pointed in a different direction.'

For the third event, it is hard to see a direct impact on the Service's PB system and the attitude towards is. For the environmental legislation of the 1970s, the broadening of the organization's policy goals and, as a result, increased reporting requirements can be imagined as an impact. Overall however no large event stands out that provides a strong explanation for the use of PI by the Forest Service. In addition it seems that the current system used had been largely in use before these events occurred (see Kaufman 1960: 113-118).

Asymmetry of power is another institutional factor that could provide an explanation for taking performance budgeting and performance management seriously. There may have been a certain influential unit or person that acted as a champion of output measurement or performance management that may have gained an extra strong position for whatever reason. A first suspect when it comes to introducing a performance budgeting agenda in an organization, are financial or planning units. There is no evidence of this from the documents and interviews that indicates that this was the case at the FS. Performance planning, reporting and budgeting are dominant as an activity but not particularly as an organizational unit.

Neither does a particular leader arise that in the memory of respondents ought to be credited with the extensive use of performance information. Obviously it helps to have a leader who is interested in performance figures and evidence based policy. In fact the current regional forester of R6 is credited with such an attitude which fits the good reputation that R6 has with regard to performance budgeting. One respondent explains:

'Our current regional forester is a visionary. He has a doctor's degree in forestry economics. He makes a difference because he gets it how it works (for example you cannot do faster forest restorations with less dollars). This helps in the communication. He supports the use of hard numbers, he likes facts.'

In explaining a service wide phenomenon, the current regional chief's favourable attitude towards performance management does obviously not provide enough of an explanation. Looking back somewhat further in the Service's history, a champion of performance measurement may be identified in the person of Gifford Pinchot. The energetic founder and first chief of the FS is generally credited with promoting scientific approach to forestry at the beginning of the 20th century. It remains unclear however to what extent the current method of performance budgeting that has at least been in place since the late 1950s, can directly be traced to his vision of how the service should be managed.

Another way in which asymmetry of power can affect an organization's emphasis on performance measurement is a dominant role of specialists in a certain policy field. This may explain why objective measurement is regarded higher than other more political factors. To determine this, respondents were asked to choose between factors that they felt were most influential to Forest service's policy in two subsequent questions:

A) Vested interests   or   B) Policy ideas

A) Politics/public opinion   or   B) Specialist expertise

If for both questions answer B is chosen, this would indicates that Forest Service's policy area is run pretty much by experts and their policy ideas. This is regarded as highly compatible with a performance based approach to management and budgeting. However, no clear picture emerged when applying this framework. The FS views itself as a specialist impartial institute that has to settle disputes over land use between different interests using scientific arguments or as one respondent put it:

'We're supposed to be a science based organization'

Many of the Service's activities go unnoticed by most and sometimes only receive the attention of specialists. In other matters, the interests of local parties are quite apparent. In fact they are felt constantly by some people in the field. This seems to be inherent to the nature of the Service's work as explained quite clearly by respondents:

'Our work is noticed more by selected and interested parties than by masses (America has become urbanized). Firefighting does get national attention. Otherwise people mostly are tuned in because they have an interest in environmental issues, recreating or they live near a forest.'

'Activities happen right there in people's literal backyard, their communities. So a direct local political pressure comes to bear at the FS compared to a non-land management agency…..Politics extremely influence the Forest Service's work but rightly so. Healthy tensions between political and social pressures are balanced with the science side of the Forest Service.'

In addition to the coping with all kinds of local interests, Washington policies shape the work of people at all levels in the organization. Not unlike the situation with the forest service in the Netherlands, this has not always been the case. As one respondent assesses the introduction of the set of environmental laws in the 1970s:

'We are out there doing our thing, than there is a national umbrella of laws about how we do our work now and into the future.'

The specialist nature of the Forest Service's work may partly explains the advancement of performance management and performance budgeting. On the other hand, the role of objective, scientific data in decision making is constantly challenged by interests and sometimes political preferences, perhaps even more so than in other public sector agencies.

Table 7.2   Presence indicators historical neo-institutionalism

Indicators		Presence in USFS case
H 1.1	Critical juncture in the accountability chain or policy field	Present to some extent
H 1.2	Problem to which PB was seen as a solution	Absent
H 2.1	An advocate or champion of PB in a powerful position	Absent
H.2.2	A policy field in which specialists are dominant.	Present to some extent

It can be concluded that the indicators identified from historical neo institutionalism only offer a modest and partly explanation for the Forest Service's standing with regard to performance management and performance budgeting. Although the influence of the Service's founder and first Chief did not come up in the interviews (under indicator H2.1), it may have been significant and may requires further study.

7.3.2   Explanations from Sociological Neo Institutionalism

From a viewpoint of sociological neo institutionalism, the ability to openly discuss problems and reflect on existing positions make for an organizational culture in which using of PI for learning is regarded as appropriate behavior. This situation can easily be contrasted with many government organizations in which bad news and self-criticism are met with suspicion and defensive routines. The conditions for organizational learning were tested among others by letting respondents react to 4 statements:

e. Within the Forest Service, important issues are being discussed openly and fairly
f. Within the Forest Service, existing opinions are regularly challenged and discussed
g. If the FS is being confronted with a problem, a thorough problem analysis takes place prior to taking action
h. If things don't work out in the Forest Service's policies and execution, lessons are usually learned

The answers to these statements portray the FS as a learning organization with an open culture in which results and improvements are discussed easily at all levels. The drawing of lessons after taking action seems to be strongly embedded in the Service's culture.

The FS spends a lot of effort on research and evaluation, arguably more than any other service thanks to the relative size of its research branch that takes up over 5% of the Service's resources. Examples of applied lessons from PI and evaluation could be mentioned fairly easily by respondents. FS employees are aware of this as is evident from these quotes from respondents:

'In my opinion, the FS has highest degree of applied science of any federal agency.'

'We do a lot of work with the R&D stations. We read their outputs and make changes. This happens continuously. We have a definite partnership with our scientists. The only time we have disagreements with them if we have to make a decision when there is a grey area and they want to study it more.'

The aspiration to be a learning organization is communicated consistently by management. On the other hand it is acknowledged that there can be some tension between the inclination to sit down and analyze what the best course of action is and the urgency in which some of the Service's tasks need to get done as these quotes from respondents illustrate:

'We talk a lot about the learning organization. We have a big push to become learning organization. We are aspiring to that. There is a large awareness of the value of being a learning organization but the press of daily business sometimes moves you on to the next task at hand rather than learning form what you just did.'

'There is a lot of good analytical work done form all scientific disciplines. Our work also requires the ability to drop what you're doing and focus on something else (fires). We often have to make choices on external influences instead of analysis'.

Perhaps surprisingly however, the strongest learning culture seems to be rooted in the fire management branch within the Service. In fire management lessons do result in changes. For example a recent national program called *Safety Journey* addressed the tragic reality of firefighters that keep dying in forest fires. In this campaign firefighters were educated about the psychology of being safe versus the *can do* attitude where firefighters are known for. This followed newly gained insights from an evaluation that showed the limited effects on prior measures that focused on beefing up rules and intensifying training. Some comments even suggest that the evaluation and learning routines from fire management have been influencing the learning culture in other areas:

'As a result form field work especially the danger form fires we often do after action reports. Some of this culture seeps out to the rest of the agency. Even in the budget and finance group we don't really have day to day dangers but when things happen we often do a review to see what went well and what can be improved.'

'I was involved in planning an orientation day for new employees. After it was finished the committee immediately got together and had this discussion on any problems that occurred. They made notes so the next time we did the same orientation, people would have a number of action reports to look into.'

Insofar there is hesitation about learning routines this primarily concerns the Service's ex ante problem analysis prior to taking action and the fact that some action is taken for political reasons disregarding the careful analysis that is done as expressed by these quotes from respondents:

'We are perceived as shooting from the hip. We are at the infant stage of doing this better.'

'There is a lot of analysis done but a lot of action is taken for political reasons and is not related to the analysis.'

It could be added that at WO level, as noted earlier, there is not always a clear sight on what causes underperformance of certain targets because knowledge from the field is lacking to do a proper analysis right away.

As encountered in the other cases, the culture at the FS too seems to encourage discussing results and addressing issues does not seem to be greatly hampered by differences in hierarchical levels or taboos on disagreement. The ability to critically reflect on their own policy choices and results was also noticed as a characteristic of the Forest Service's culture

by Kaufman when writing about the accuracy of reporting by forest rangers and is affirmed by respondent's comments today:

..Reporting tends to be highly accurate, and field officers turn in information that sometimes reveals their own weaknesses and mistakes as well as their competence and their triumphs. (Kaufman 1960:130):

> *'We have a culture where it seems to be okay to disagree with one another. We have a lot of scientific disciplines working for the same organization. Obviously there is disagreement about what people think that the priority is for a given landscape. It is not seen at disharmonious not to agree on anything. So people feel comfortable to say that they don't agree.'*

An example of the way of a constructive internal performance dialogue is the State of the Region event (see Chapter 7.2.2) where meeting or exceeding targets is celebrated and failure to meet targets is discussed jointly to address improvement measures.

It is apparently risky to talk about the cultural openness of an organization as large and geographically dispersed as the FS as there will be and are differences between the different units and levels within the organization.

The openness of the Forestry Service's culture does not mean that there aren't subjects that are more sensitive than others. In that sense the participative openness seems somewhat higher than reflective openness. This goes for example for the administrative burden and for contradictions between the old versus the new culture that sometime pop up as these quotes from some respondents show:

> *'Our organization tends to learn and progresses based on knowledge. For example we won't even think about doing the approach to logging we did in the 1970s. But it is hard to question the red tape or change direction. We are still politically driven.'*

> *'As an agency we can be somewhat nostalgic. There is a lot of history, pride and tradition. It is OK to celebrate that but if you look back too much you you're not looking at the future.'*

> *'We do continually challenge our way of thinking and we continually bud up against the old thinking versus the new way.'*

The matter of old versus new culture within the FS was already briefly mentioned in CH 7.1. Although somewhat intangible, the old culture seems to refer to the nostalgic image of the days when the Service was mainly about generating timber sales and the ideal forest ranger had an image of a masculine hero facing rough conditions in a remote area. What the new

culture exactly looks like is still remains somewhat unclear. One respondent reflected on the changes the Service went through in the last couple of decades:

> 'We have been going through a change for a while now. When I began working for the Forest Service nearly 40 years ago, we made money for the Treasury through the volume of our timber sale receipts. We've been moving away from that since the 1990's and it is changing our culture. Although this began over 20 years ago, it takes a long time in any culture to completely move to something different. The old culture of just timber as our total focus has shifted to forest health and later to landscape restoration. When our timber program was reduced our workforce was reduced; primarily in smaller communities where the majority of work is accomplished. Our workforce and budget have both been reduced by 50 percent over the past 20 years. This is changing the way we work. In addition to the budget and workforce reductions, technology has also changed the way people do their work. Information is much more readily available. That changes our work and the way we think about things.'

An organizational culture can either help or hamper the use of PI as it may heavily influence the way (performance) information is selected, interpreted, presented and processed. If there is a joint understanding across organization members about the usefulness or even necessity of performance data and this interpretation is shared by its principal(s) and other stakeholders, measurement and reporting these data is will be considered an undisputed part of one's job. This is expected to be beneficial to using this information to learn and improve. Respondent's comments suggest that this may play a role at the FS too when it comes to performance measurement and reporting:

> 'We know the reason why. I think we understand why Congress gives us certain money in certain arenas.'

> 'When Congress passes an appropriation for the FS, they don't just pass the dollars, they pass the target. Our chief testifies for what he is going to achieve as a return for these dollars. Historically this is what we do, we are expected to perform. Some benefits are less measurable. Contacts with the communities not handled so much by budget but more by personnel.'

> 'There is a pride in the FS. People pride themselves on meeting their targets and do a tremendous amount of to meet them. It can send a mixed message to the Hill though when targets are met anyway in times of budget cuts.'

People generally seem to be convinced of the usefulness of measuring and reporting PI for the sake of improving effectiveness as well as for accountability as one respondent summarizes quite clearly:

'Because it is fed by local data and you get aggregated reporting, like for instance on watershed health and species diversity, it helps in the long term. In the short term as a manager it is often hard to see that benefit. If we weren't doing that we wouldn't be as accountable to taxpayers as we should be.'

That does not mean that every quantitative target is seen as equally useful. The extensive reporting requirements with in areas where there are soft targets apply and monitoring and results accountability is less rigid are questioned by some:

'There's a whole bunch them that feel almost most like busy work. In recreation they would do al off these things if they didn't have the targets. Like the formula that uses persons visiting recreation sites. Does that help us see that we should have more or less people come? This would happen regardless of whether there was anything written down about numbers.'

In conclusion however, the evidence found portrays the FS as a learning organization with an open culture in which results and improvements are discussed easily at all levels. Drawing lessons after taking action seems to be strongly embedded in the Service's culture. Perhaps surprisingly, the activities that most require a prompt response (fire), are characterized by organization members as the field with the strongest learning routines. Performance measurement and reporting seems to be a rather undisputed part of the work of almost all organization members and is generally perceived as useful. At the same time also some tension remains with regard to discussing differences between the old and the new culture, administrative red tape.

Table 7.3   Indicators from sociological neo-institutionalism

Indicators		*Presence in USFS case*
S 1.1	Absence learning disabilities	Clearly present
S 1.2	Participative openness	Clearly present
S 1.3	Reflective openness	Present to some extent
S 2.1	Shared view of the meaning of measured performance	Clearly present

7.4   Contextual factors

It is hard to ignore the difficult budgetary situation part of the time this case study was conducted as the federal government was running under a sequester meaning that all agencies were facing automatic spending cuts. This exceptional situation started after Congress failed to agree on budget resolution as well as on extending the nation's debt

ceiling. Although it is interesting to think about the implications of a government running on systematically reduced inputs for the entire concept of goal based performance budgeting, the implications for the FS system of performance budgeting and management were not apparent. Although the sequestration certainly affects targets, when asked whether the sequester impacted FS strategy a respondent reacted that strategy states that in what direction you want to move and not necessarily in what pace you need to get there.

Another development that feeds worries about the Service not being able to live up to future expectations in terms of performance levels are the rising costs of forest fires. As money necessarily is being shifted form fire prevention to fire suppression each year and total budget keeps shrinking, something will have to give at some point. This is a development that has been going on for quite some time however.

Eventful as these developments may be they do not provide an apparent explanation for the level of PI use by the FS. In fact one can argue that the FS system of linking of outputs to BLI funding and the robust methods of budgeting in the field (see Box 7.1) prepare the Service relatively well for this kind of budgetary crisis. On the other hand one could imagine that a consistent system like the FS output based funding system may become increasingly vulnerable to political pressure in an environment of shrinking budgets and ad hoc priority setting.

7.5   Conclusion

The investigation of the indicators for the FS should provide answers to three questions to be answered for each case (see CH 4):

A. *To what extent is this a PB success?*

B *Are the identified conditions present?*

C. *Does context offer a likely explanation?*

The answer to the first two questions can be given when looking at the numerical scores, obtained from the interviews and expressed in the range of 0-2 and applied to the different indicators (see Figure 7.7).

Figure 7.7   Numerical scores of the indicators from the interviews
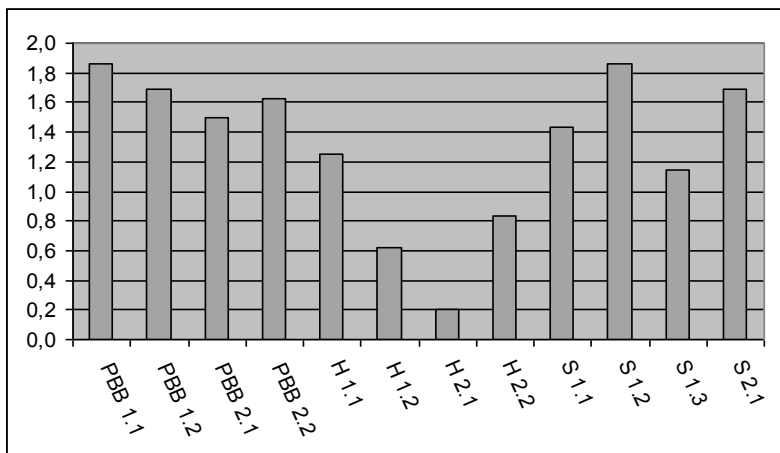
The indicators for PB show that the FS easily fits all the criteria identified for being a PB success. Performance budgeting is a pivotal element of budget formulation and plays a clear role in controlling activities undertaken by officers in the field working in the regions. Alignment of all goals with budgets and outputs is consistent but complex and may appear somewhat fragmented. The PB framework is well designed and *de facto* use is convincing though not equally so for all policy goals.

The second question should be answered with a yes for the sociological neo institutional indicators. The FS fits the criteria for a learning organization, has an open culture with respect to discussing results and organization members seem to share a positive perception of the usefulness of measured performance data. The ability to critically self-reflect and innovate is there but can sometimes be bounded by nostalgic elements of the Service's culture. There was little evidence on a strong presence of the indicators from historical neo-institutionalism. Some events may have intensified a certain development of the PB system but this system was in essence already when these events occurred.

Contextual factors do not offer a likely explanation for the PB success found in this case although they could challenge continuation of this success in the future.

Epilogue
The variation in size and landscape and the remoteness of units to each other and the headquarters are clearly greater issues in the US case but were recognized as organizational challenges in the SBB case as well. But the similarities do not stop there. Intrinsic motivation and professional pride of FS employees matches those for SBB and may provide a logical explanation for the result orientation and purposeful PI use found throughout the organization. It has been suggested by some that PI use and performance positively correlate with intrinsic motivation of employees for the organization's tasks or public service

motivation as it is often fashionably referred to these days (Perry&Hondeghem 2008:8, Moynihan & Pandey 2010: 11).

Although the historical institutional indicators tested in this study do not score as plausible explanations, the FS' beneficial culture towards PB can only be explained by looking at historical institutional factors. As Herbert Kaufman's explains in his late 1950s study of the Forest Service (a work thankfully used for this study), the FS has traditionally relied heavily on training and indoctrination as a tools to of integration to counter the apparent tendencies for organizational fragmentation. This resulted in a strong cultural and professional identity and a strong sense of mission of the FS. Wilson noted that a dominant single culture or sense of mission can be beneficial to the flow of information within government agencies:

When a single culture is broadly shared and warmly endorsed it is a mission. The great advantage of a mission is that it permits the head of the agency to be more confident that operators will act in particular ways that the head would have acted had he or she been in their shoes. There are fewer distortions in the flow of information because sender and recipient of the message share common understanding. (Wilson 1989:109)

In the late 1980s Wilson even regarded the US Forest Service, along with the FBI and the Army Corps of Engineers, as a prestigious elite agency thanks to their strong sense of mission and highly developed professional culture. It is therefore somewhat puzzling to find that in the 2012 employee satisfaction ranking all these three agencies rank well below average (http://bestplacestowork.org/BPTW/rankings). This raises the question if a professional culture is that stable a characteristic from an institutional point of view.

The strong professional culture and the highly standardized administrative procedures (including accomplishment reporting) makes for an organization where employees seem to find performance measurement and reporting to be a self-evident part of their job. In this respect Kaufman already reported a strong administrative reporting culture reinforced by technical development (Kaufman 1960: 127-9). Filling the extensive administrative system is indeed still an important part of a forest ranger's job. Technological development is easing this task somewhat but increasing precision and a higher demand for accountability is requiring additional time as well. This was already described by Kaufman but still is true of today's FS:

IBM machines, for example have been installed in all the regional offices, and virtually all bookkeeping and accounting are performed on them. Data coming in from the field units are punched on cards and processed by regional office technicians, and the printed figures are distributed to the relevant officials in the regional offices, to the appropriate forest supervisors, and, by the latter, to the Rangers. A glance at running totals each month informs everyone what is happening on every district, and as a result, most Rangers, at one time or another, have had their attention called pointedly to some trend or imbalance requiring

redress. The load on field men is eased by the service, it is true, but the service also improves the detection of deviation from promulgated norms. (Kaufman 1960:134-5)

> *'Technical advancement also has a downside: a lot more analysis is required by the public that increasingly questions what we do.'* (respondent in 2013)

Another similarity with the other case studies is that the essence of the PB system was developed long ago within the organization and was gradually applied more broadly to other organizational goals, products, and processes. In the case of the FS, performance planning and accountability clearly grew from the timber program as one respondent explained:

> *'From the very start there was interest in how much timber was produced and there has always been some level accountability. Our precision has gotten a lot better but the concept has not really changed.'*

The gradual bottom up development of a PB system contrasts the reality of many public organizations that adopted a PB system which is built and adopted quite suddenly in response to a principal's needs regarding goal alignment and result accountability or because of top down government wide regulations.

The image of the wide array of objectives and measures being 'poured down' on a forester and his or her unit (see Figure 7.2), conflicts with the classical notion of organizational strategy portrayed as a pyramid in which overarching goals are broken down into partial tasks executed by the lower levels. Instead one can view policy execution at the Forest Service as well as the other cases studies as an inversed pyramid, or a funnel as Kaufman named it, with the forester at the lowest level having to make the difficult judgment between conflicting strategic goals that are demanded from his or her organization. An illustration of such a dilemma in this case was a decision to open up a forest road to 4x4 vehicles. This would serve the President's agenda of increasing the forest's attractiveness to city dwellers but at the same time may increase vulnerability of nearby habitat of endangered species.

The complexity of strategic and performance planning has proved to be an obstacle for linking funding to performance as was highlighted by several reports of the Government Accountability Office (GAO). On the one hand It would seem a fair assessment to state that the Service's stable annual system of output funding and measurement has been able to accommodate and support increasing requirements of performance planning in the late 1990s and political prioritization of recent years to a great degree. On the other it may be argued that the Service's tradition reliance on output reporting and linking inputs to measured outputs slowed the accommodation of a more modern performance budgeting system that centers programs, policy outcomes and budgetary flexibility. As discussed, the traditionally strong professional culture of the FS traditionally is seen as a favorable circumstance for successful PB but may also have a downside as some respondents indicated

that they felt that innovation at times conflicted with the service's nostalgic culture. In that sense the Forest Service, to some degree, may become the victim of the principle of the *Law of the handicap of a head start*[18] in recent years.

Finally, although not the primary focus of this study, the FS cases study also illustrated how a PB system can function under pressure from a chaotic input oriented budgeting process as we have recently seen in DC. The FS budgetary process seems reasonably well prepared for dealing with gradual shrinking of budgets and continuous uncertainty about authorized allocation. On the other hand the FS has also profited from temporary extra funding form the federal recovery program (ARRA). Due to years of underfunding actual targets, this money was spent at the forest level more or less in line with the metaphor the squeaky wheel[19] as a respondent indicated:

> *'ARRA money was used to do some long overdue maintenance of the remaining volcano visitor center. But it didn't really affect our overall facilities program.'*

> *'Priorities depend on leader's focus areas. This depends on who's the squeaky wheel at the time but a lot is also dependent upon what is knocking on the political door.'*

Squeaky wheel budgeting represents a fundamental different approach than the rational budgeting process based on realistic cost prices and workloads as developed by the FS throughout the years. As such it raises the question to what extent a well-developed PB system can be compatible with continuous reallocation process that characterizes a long period of fiscal consolidation

Data collection USFS case
The primary method of data collection were semi structured interviews. In September/October 2012 and June/July 2013, ten persons were interviewed with regard to the USFS case. Two of these worked at the President's Office of Management and Budget charged with oversight on performance on key priorities and budgetary issues. At the FS Washington Office the branch Chief Performance Management was interviewed twice. At the regional HQ of Region 6 in Portland, Oregon three persons were interviewed representing budget and performance staff and program management. Within R6, 4 persons were interviewed at the Gifford Pinchot National Forest representing line management, budget & performance staff and program management. Data were primarily processed by qualitative analysis although some quantitative analysis was used for comparing and aggregating results (see section 4.6). These scores were aggregated for each case allowing

---

[18] The **Law of the handicap of a head start** (original Dutch: *Wet van de remmende voorsprong*) or **dialectics of lead** is a theory that suggests that getting an initial head start in a given area may result in being a handicap in the long-term. The term was coined in 1937 by Jan Romein, a Dutch journalist and historian, in his essay "The dialectics of progress" ("De dialectiek van de vooruitgang"),
[19] The expression *'the squeeky wheel gets the grease'* refers to the notion that the most noticeable (or loudest) problems are the ones most likely to get attention

for a qualitative comparison on an ordinal scale. After finishing each draft case study, the report was reviewed by a key informant from the organization involved. By means of triangulation the findings from interviews were compared with anonymous questionnaire results and findings from analysis of documents sometimes nuancing the findings from the interviews. Appendix III shows the questionnaire that was used.

Sources

Absher, James D. 2008, *Partnerships and Volunteers in the U.S. Forest Service*
Paper prepared for Proceedings of the 2008 Northeastern Recreation Research Symposium

Department of Agriculture, Office of Inspector General, Audit Report: Forest Service's Financial Statements for Fiscal Years 2010 and 2009, 08401-11-FM (Washington, D.C., November 2010).

GAO-03-503, May 1, 2003

GAO-09-443T

GAO-11-423T, Mar 10, 2011

USDA FS Strategic Plan 2007-2012

USDA fy2013 budget justification

USDA fy2013 budget overview

USDA fy2014 budget overview

USDA fy2007 budget overview

USDA fy2010 budget overview

www.fs.fed.us

CHAPTER 8 CASE STUDY U.S. FEDERAL AVIATION ADMINISTRATION -AIR TRAFFIC CONTROL ORGANIZATION

8.1   Description of the Agency and its Principal

History and tasks FAA/ATO

The Federal Aviation Administration is a federal agency that resides under the U.S. Department of Transport (DOT). The FAA is tasked with overseeing U.S. civil aviation in all of its aspect as expressed by the agency's mission: *to provide the safest, most efficient aerospace system in the world*. To achieve this, the FAA establishes and enforces aviation regulations, oversees inspections and certifications and maintains the integrity and reliability of the civil aviation infrastructure. However, in terms of people and budget, providing air traffic control services forms the largest part of FAA's work. On a daily basis some 15.000 air traffic controllers handle around 100.000 flights in U.S. airspace, carrying about 2.5 million passengers. On average, that comes down to one take-off and one landing every second somewhere in the US.

The history of the FAA goes back to the 1920s when the first aviation related legislation was passed (the Air Mail Act, followed by the Air Commerce Act). In 1958, Congress created an independent federal aviation agency and in 1966, the Department of Transportation was created, with the Federal Aviation Administration becoming one of the modes within DOT. Air traffic control operations began in 1936 with the Bureau of Air Commerce started operating 3 traffic control centers that employed 15 controllers in total. In December 2000, a presidential executive order directed FAA to create a performance-based organization that focused solely on efficient operation of the air traffic control system. The new organization was to be headed by a Chief Operating Officer and by 2003 Russell Chew was selected to become first COO of the newly formed Air Traffic Organization (ATO). The ATO consolidated FAA's air traffic services, research and acquisitions, and Free Flight Program activities into a smaller, more efficient organization with a strict focus on providing the best service for the best value to the aviation industry and the traveling public. The ATO officially began operations on February 8, 2004.

Unlike its counterparts in many countries (including LVNL in the Netherlands), FAA/ATO was not separated from the transport ministry to operate as an agency at arm's length from the Transport department funded directly by its customers. Actually, in the 1990s the Clinton administration, as part of Vice President Al Gore's "reinventing government" initiative, did propose to do just that but this proposal was never adopted.

Size

The FAA currently employs close to 49,000 people (47,031 FTE in FY12). About 4,000 FAA employees work at FAA Headquarters in Washington D.C. Another 1,500 work at the

Technical Center in New Jersey and 3,500 work at the Aeronautical Center in Oklahoma. The remaining 40,000 people work at regional and field offices. The bulk of them are ATO personnel (33,315 FTE in FY12), located at airports across the US. ATO operates a total of 21 En Route Air Traffic Control Centers, 166 terminal radar control facilities (TRACONs), and another 380 air traffic control towers operated by the FAA or contracted controllers. Besides the 15,000 air traffic controllers, the ATO's workforce is made up by the technicians, engineers and support personnel. Around 9,000 technical employees make sure that more than 41,000 pieces of equipment operate daily.

FAA's annual budget is around $16 billion (enacted FY2011). About 70% of this is used to fund the ATO. Airports (ARP) make up another 20% of budget expenses. The FAA's collects about $ 0.6 billion in revenues annually, mainly from service fees (e.g., landing, registry, and overflight fees) and through reimbursements for products and services provided to domestic and foreign governmental entities.

Organization

The FAA is led by a politically appointed administrator who heads a rather complex organizational structure. Although many of its tasks are performed across the many locations throughout the US, most major units reside in Washington D.C. Budgeting and performance issues are also handled primarily from the central headquarters. The FAA's main tasks are divided into four so called 'lines of business':

- Air Traffic Organization (ATO): provides air traffic control operations for commercial private and (to some extent) military aviation.
- Aviation Safety (AVS): oversees safety of aircraft, competency of pilots and mechanics and provides safety rules and standards
- Airports (ARP): provides grants to state and local governments to improve airport safety efficiency and environmental quality in addition to standard setting and inspection.
- Commercial Space Transportation (AST): regulates the safety of U.S. commercial space transportation. This is a relative new line of business that gained importance after retirement of NASA's Space Shuttle. The two main challenges for AST today consist of getting crewmembers to and from the International Space Station and regulating the growing potential for space tourism.

The four lines of business correspond with four organizational units that are each lead by an associate administrator. In addition to these four units there are another 9 unit that provide central services at the headquarters such as communications, human resources and finance. These are headed by an assistant administrator. As mentioned, the ATO is by far the largest FAA unit. Eight service units make up the ATO. Of these, Air Traffic Services and Technical Operations are the most numerous in terms of personnel (combined these units employ over 90% of ATO personnel). Management Services is tasked with maintaining and managing

the performance management systems and inform management to make sure that performance stays on track. Within ATO's core activity of air traffic control different types of facilities and functions can be distinguished as illustrated in Figure 8.1.

Figure 8.1   Types of facilities and functions within air traffic control (FAA Workforce plan 2012-2021)
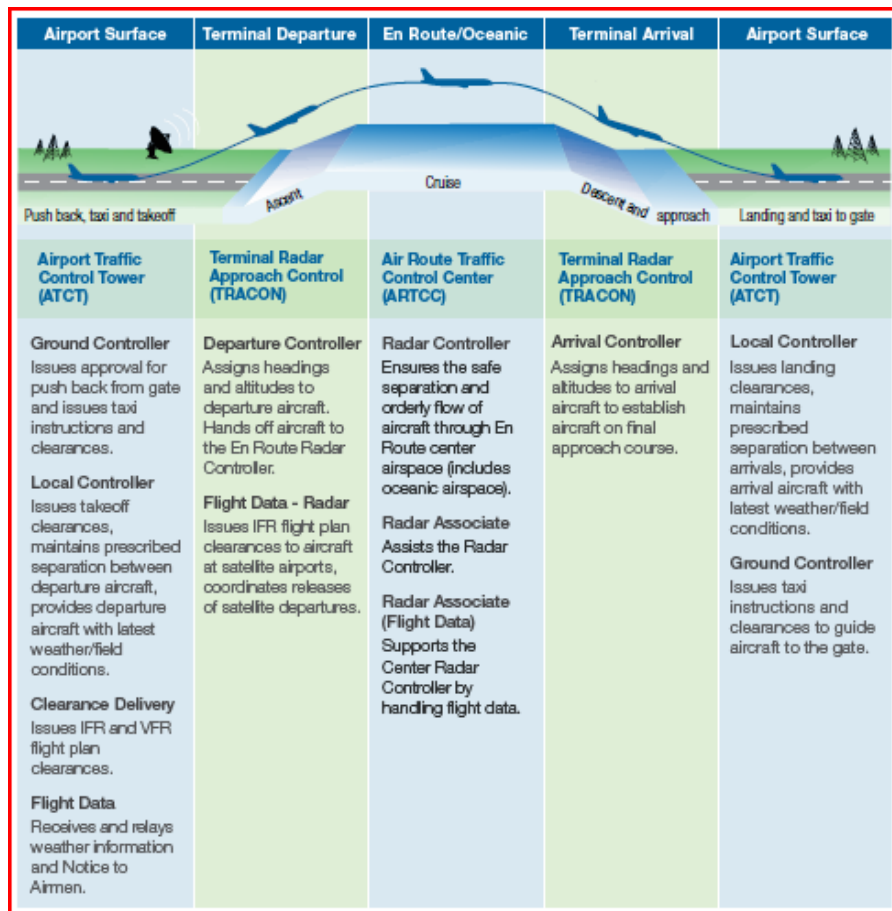
Regionally the National Airspace that FAA manages is divided into nine regions with a regional office each (see Figure 8.1). The region oversees local airport development, certification, and safety. Air traffic control operations and performance planning and monitoring however are highly centralized and are managed directly from the DC headquarters.

Figure 8.2   FAA's nine airspace regions and regional FAA offices

Policy field and tasks principal

The FAA is one of thirteen agencies that are part of the Department of Transportation (DOT). As such the FAA is responsible for achieving the strategic goals DOT is set to achieve. The DOT strategic plan currently applicable has been the DOT 2012-2016 Strategic Plan named 'Transportation for a New Generation'. This plan contains five strategic goals, four of which touch (partly) upon the FAA's responsibilities:

- Safety: *Improve public health and safety by reducing transportation related fatalities and injuries*. This is translated into 14 performance measures, 3 of which are FAA related:
  - Reduce commercial aviation air carrier fatalities to no more than 7.4 per 100 million persons on board
  - Reduce general aviation fatal accident rate per 100.000 flt hrs to no more than 1.06
  - Reduce category A&B runway incursions to no more than 0.395 per million operations
- State of good repair: *Ensure the U.S. proactively maintains critical transportation infrastructure in a state of good repair*. This is translated into 5 performance measures, 1 of which is FAA related:
  - 93% of paved runways should be in excellent, good or fair condition.
- Economic Competitiveness: *Promote transportation policies and investments that bring lasting and equitable economic benefits to the nation and its citizens*. This is translated into 13 performance measures, 3 of which are FAA related:

- Maintain average daily airport capacity for Core Airports of 86,835 arrivals and departures.
- Maintain operational availability of National Airspace System (NAS) at 99.7%
- Maintain NAS on-time arrival rate at Core airports at 88%

In addition a core goal without quantitative metrics is formulated:
- replace a 40 year old computer system in 20 air traffic control centers with a new one (ERAM)

- Environmental Sustainability: *Advance environmentally sustainable policies and investments that reduce carbon and other harmful emissions from transportation sources*. This is translated into 11 performance measures, 2 of which are FAA related:
  - Improve National Airspace System energy efficiency by at least 2% annually (from 4.24 Tg/Bkm in 2010 to 3.73 in 2016)
  - Improve Aviation Noise Exposure from 307,420 persons in 2011 by at least 2% annually to 328,000 in 2016.

A fifth strategic goal, organizational excellence, does affect FAA but is not translated into quantitative performance targets. This goal includes a number of initiatives to support workforce diversification, emergency preparedness, open government and financial performance. DOT's sixth and final strategic goal strategic goal, liveable communities, concerns and involves increasing accessibility and transportation choices of communities and does not directly involve FAA targets and activities.

So total of 10 quite specific DOT targets have to be realized by the FAA. The current DOT strategic plan falls in line with a tradition of quite specific quantitative target setting by the department and its major agencies. In aviation as well as other transport policy areas such as road safety, safety improvement has historically leaned heavily on quantitative measurement, analysis and evaluation.

This has delivered spectacular results as illustrated by this statement by Thomas Hendricks, senior vice president for operations and safety of Airlines for America (Bloomberg 2012):

'The risk of a fatal accident in commercial aviation has been reduced to 1 out of 49 million flights over the past five years, from 1 in 1.7 million flights from 1975 to 1989, according to NTSB records. That's a 96 percent decrease in risk. Safety has improved since the late 1990s as the airline industry and regulators learned to analyze massive quantities of data for anomalies and voluntarily made changes to head off potential problems'

Choice of principal and agent
Like any other policy area, different principal-agent constellations can be chosen to assess the degree of PB implementation between organizational entities. In this case the ATO was chosen as the primary focus for the position of the agent. The reason for this is that ATO represents the dominant operational unit within FAA where most primary processes and

activities take place that are expected to fulfill the policy objectives and realize targets. In addition the ATO seems to share complete commonality in type of tasks with LVNL, the subject of an earlier case study. FAA as a whole makes for a less fit with LNVL because it is responsible for a broader array of tasks such as airport regulation, airplane inspection and pilot certification.

Figure 8.3   Different principal-agent relationships
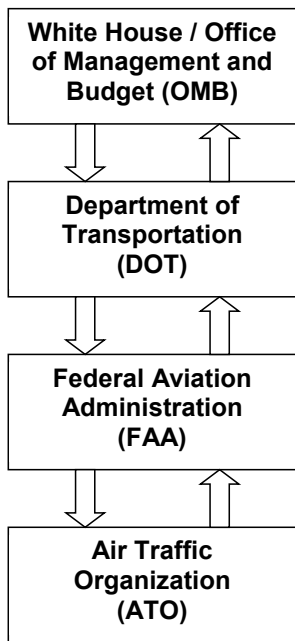
When determining the ATO's principal, several organizations qualify as illustrated in Figure 8.3. First there is FAA where the ATO formally resides under as a unit. Secondly there is DOT of which FAA the largest residing agency. DOT officially mandates performance targets to FAA and the FAA budget request is part of DOT's budget request. Finally there is the White House's Office of Management and Budget (OMB) with whom DOT negotiates budget and performance targets and to whom DOT has to report its performance. When it comes to accountability and performance dialogue however these lines seem to be more blurred than they occur in Figure 8.3. The distinction between FAA and the ATO is not always easy to make in practice as the ATO forms the operational backbone of the FAA. As a result, the ATO's performance targets are the subject of discussions by FAA management on a daily basis. The unclear distinction between the different principals to the ATO is not a too rigid one in practice is also illustrated by responses form the interviews with DOT and FAA employees (in that order below):

'For the most part they (FAA/ATO) take whatever information they have already generated…. and they just transfer it over into the departmental plan. To them it is not a big exercise. For the most part we adopt what they come up with' (DOT)

'Because we are a big part of DOT and we are very visible, we have our own OMB people assigned directly and not through DOT, we keep them informed.'

Nonetheless, DOT was chosen as the main principal to ATO in this case study because the set of targets mandated by DOT mainly apply to the operations as conducted by ATO. In

addition there is a direct performance dialogue between ATO leadership and DOT as well as with OMB.

Obviously Figure 8.3 could be made even more complete when including decentralized units answering to ATO headquarters in DC. These were not included because of the relative low level of discretion of these type of units (e.g. compared to the U.S. Forest Service case study)

8.2   Degree of PB implementation

The FAA has long been considered a leader among agencies when it comes to implementing data driven, performance management in the US federal government (Mills 2013: 27, IBM 2011:2). In fact, when DOT was required to meet quarterly over performance issues under the revised federal performance management regulations issues in 2010 (GPRAMA), the ATO already had a weekly meeting with the DOT deputy secretary on performance issues making the new requirement seem somewhat redundant according to some of the staff involved.

When it comes to the performance budgeting aspect, the FAA, like any other federal agency, is funded on the basis of line items. This traditional way of input funding provides a reasonable degree of stability but means that funding levels have to be aligned annually with productivity levels and the performance targets from departmental and agency strategic plans.

The FAA is funded by four different appropriations (or budget line items):
- Operations –This is FAA's largest appropriation and finances the air traffic control and air navigation systems.
- Grants-in-Aid for Airports (AIP) – funds various programs to improve airports
- Facilities and Equipment (F&E) – funds modernizing and improving ATC and airway facilities, notably the NextGen program
- Research, Engineering, and Development (R,E&D) – funds all kinds of aviation related research (e.g. propulsion, materials, safety, environmental, weather)

The three capital investment appropriations (AIP, F&E and R,E&D) as well as about half of the Operations appropriation are funded by the Airport and Airway Trust Fund (AATF) that receives its revenues from taxes to airspace users. Therefore it is fair to say that, depending on the year, some 70-80% of FAA budget is funded indirectly by their clients, the airspace users.

Figure 8.4   Relative allocation of FAA appropriations
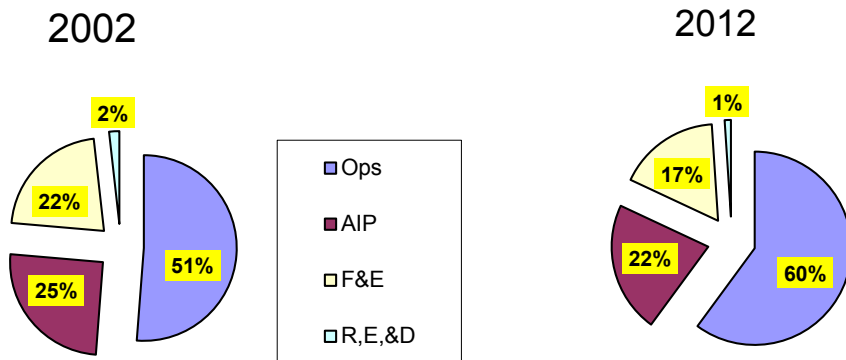
As illustrated by figure 8.2 the relative share of FAA's budget that goes to operations differs significantly between 2002 and 2012. Although the costs of operations have been growing steadily over the years in absolute terms, this percentage fluctuated roughly between 50 and 60% during the 2002-2012 period.

As personnel compensation makes up roughly half of FAA's budget, the outcome of the budgetary process is largely dependent upon negotiations with the powerful unions. The investment budget is the more discretionary part of the budget although earmarked for capital investment by appropriation by three separate line items. The relationship between funding and performance therefore appears rather indirect as expressed by one of the FAA respondents:

> '*Some measures are more tied to funding than others. For safety, capacity, efficiency we basically track how well aircraft are moving through our system. There is not a lot of funding that directly supports these metrics.*'

On the other hand, the FAA is the only federal agency that has had a consistent and large scale pay for performance system in place. That indicates that serious attempts have been made to link the largest component of the budget, namely personnel, to individual and collective financial incentives. Section 8.2.2 will go further into the whereabouts of this system.

8.2.1 PB in the relationship DOT – FAA

Just like the 2012-2016 Strategic Plan is the current strategic plan for DOT, the 2009-2013 FAA Flight Plan has been the main official document concerning performance planning for FAA and therefore ATO. The plan contains four strategic goals:

- Increased Safety, translated into 6 objectives and 8 performance targets
- Greater Capacity, translated into 3 objectives and 7 performance targets
- International Leadership, translated into 2 objectives and 4 performance targets
- Organizational Excellence, translated into 5 objectives and 12 performance targets

As an agency within DOT, the FAA is not required to issue a separate Performance and Accountability Report (PAR) but does so anyway. Up until 2011 FAA reported on the goals and targets from its 2009-2013 Flight Plan in its PAR. The performance planning in this document was based on the previous 2006-2011 DOT strategic plan.

From 2012 however the FAA based its PAR on a new FAA strategic document named Destination 2025 (D2025) instead. This vision document as it is referred to, has five strategic goals that are translated into another 24 performance targets:

- Next Level of Safety
- Workplace of Choice
- Delivering Aviation Access through Innovation
- Sustaining our Future
- Improved Global Performance through Collaboration

Destination 2025 incorporates elements of an ambitious vision on a major overhaul of the National Airspace referred to as NextGen. NextGen revolves around the transition from radar- navigation to satellite based navigation resulting in greater airspace capacity, lower costs to airlines as well as drastic reductions in fuel emissions and noise levels. Although there are some overlapping safety targets with the DOT strategic plan, DOT acknowledges that FAA's D2025 goals are not totally aligned with theirs. At the same time the issue in terms of goal alignment was downplayed by a DOT respondent:

'Their goals are different from the Departmental goals. Their 2025 plan differs from our plan: Their goals are more focused on delivering deliverables than the outcomes per se. Although they are slightly different, it is fairly easy to translate the data. There is no issue in terms of alignment. It's a manageable tension. There are different audiences for both plans'.

For FAA and ATO this means that they currently report on different strategic planning documents at the same time, the DOT Strategic plan 2012-16 and D2025. All performance measures in the DOT Strategic plan 2012-16 appear in the FAA annual business plan for 2013. For each one of these an elaborate definition and explanation is provided (including

data source, public benefit, external factors, reliability). In addition, goals from both plans as well as targets from Flight Plan 2009-2013 are integrated. The ATO Business Plan 2013 is a subset of the FAA business plan, and is a 200 page document that specifies activities and targets. ATO is involved in all five DOT strategic goals (see section 8.1) but bears most direct responsibility for the targets of the safety and economic competitiveness goals. As a result of the plural character of strategic planning, the goal alignment between the different planning documents of DOT and FAA from a multi-year perspective appears rather complex and somewhat fragmented as is illustrated in Figure 8.5.

Figure 8.5   Strategic planning DOT, FAA and ATO
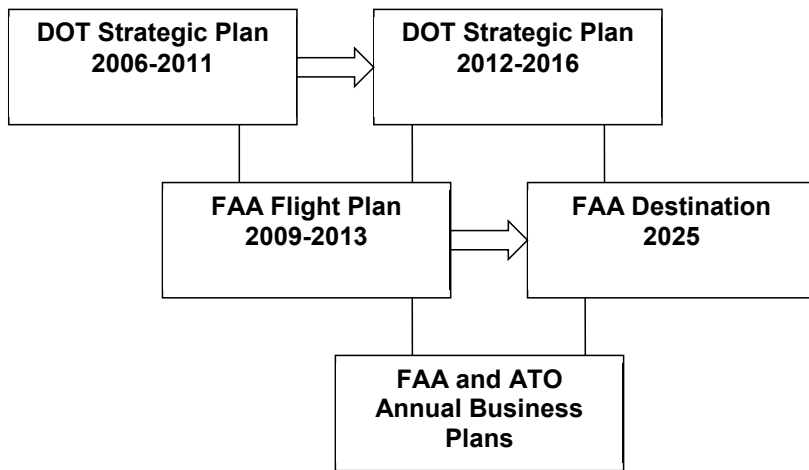
Financial reporting is in line with the DOT guidance framework and FAA expenses are all attributed to DOT strategic goals. That is why DOT is able to present the FAA expenditure attributed to their strategic goals as illustrated for FY2012 in Table 8.1:

Table 8.1   FAA spending FY2012 by DOT Strategic Goal

The budget preparation process at FAA follows guidance from DOT and managers in charge of each of FAA's four appropriations will go out to the program managers in the offices to find out what their funding needs are. This is a process that goes back and forth constantly until the request is formulated and passed on to DOT to review. Following this there are several meetings with OMB and DOT policy and budget people who are assigned to FAA. This usually takes most of the summer months and results in the request that goes to OMB as part of DOT's budget request.

In every budget submission there also is a performance budget that goes along with the budget submission. One respondent described the FAA performance budget as follows:

'The performance budget tries to weave together a narrative treatment with dollars and exhibits like various Excel worksheets and the contribution to DOT's outcomes.'

Essentially, FAA prepares the aviation portion of the DOT annual performance plan. Because FAA has its own strategic plan in addition to being part of DOT's strategic plan, the targets and activities in the internal FAA annual business plans have to be aligned with both plans. One FAA respondents explains:

'In their business plans everyone has to line up their plans with both Destination 2025 and the DOT strategic plan. After the strategic outcomes are aligned, we slice and dice them. Then we tie it with the dollars and the activities'.

ATO represents 82% of the FAA budget and the pay to air traffic controllers represents its largest cost component. These costs are allocated to goals by the type of controllers; en-route controllers to the safety goals, and terminal controllers to the capacity goals.

Actual use of PI in relationship with principals
The area of air traffic control comes across as a highly technical and specialized field of work where performance, results and funding seem hard to tell apart for a principal, let alone a Congressman or taxpayer. Despite this a critical dialogue between OMB and the FAA is reported on precisely these issues, especially during budget preparation and capital investments as one respondent explains:

'Despite the high level of technical details, we have a couple of OMB people assigned to us and scrutinize us on the relationship between funding and performance'.

'When we talk to OMB we have to provide and quantify what we are going to do with the money. OMB comes back and says what if we only give you this much what are you going to cut. I think performance and money is pretty integrated in that sense'.

Another example of PI-use in relation to funding is capital investments. According to OMB regulations, these require a business case with documentation that associate the investment with the impact on a performance measure like for example runway incursions.

Congress however operates at significant distance from ATO's work and mainly interferes when something directly affects their district such as a proposal to cut down on traffic control or close down a tower at a small airfield with little traffic. More in general it can be said that among FAA's lines of Business, airports receive most attention from Congress. When there is a meeting with Congress staff, OMB usually accompanies the FAA delegation.

Explaining FAA performance to Congress has not been easy as was illustrated in recent years when Congress scrutinized FAA after being alarmed with an increased number of reported operational errors by air traffic controllers. In fact the increased number of reported errors was the result of improvements FAA made in its incident reporting. This was viewed by FAA as good news because these incidents previously went undetected and provided valuable opportunities to learn. Besides, the number of serious incidents continued to diminish throughout the years. In an excellent analysis of this case, Mills suggests that the increased public scrutiny from stakeholders like Congress, OMB, GAO and the DOT IG may have lead FAA/ATO to rush into adopting alternative indicators that they cannot fully analyze and synthesize yet (Mills 2013: 32).

Apart from coordination during budget preparation and during contacts with Congress and OMB, a formalized continuous performance dialogue takes place between FAA and DOT throughout the year. ATO management has a weekly meeting with the DOT deputy secretary on performance issues. In addition, the FAA administrator meets with the deputy secretary on a more or less quarterly basis. One of the issues they discuss is a scorecard at the departmental level that includes measures and targets from the DOT strategic plan. On the part of DOT, the attitude towards FAA and ATO when it comes to performance reporting and monitoring seems to be quite trustworthy and relaxed and usually does involve financial re-allocation as expressed by a respondent:

'For the most part we adopt what they come up with. There are going to be a couple of instances where we're adding new things in our strategic plan like management objectives (mostly financial and HR) that they haven't done before and we need to obtain new data from them but it's not really a challenge for them. I don't worry about it too much'.

'It's more of a check because there is only so much the department can do. The administrator can ask for information to be put on a website or help the agency deal with outside partners like congressmen but there is no moving around of money or something'.

8.2.2 Performance management within FAA

As mentioned earlier, the main strategic performance planning document within FAA is currently the Destination 2025 plan. The 24 high profile measures from D2025 are more of a strategic nature and are monitored closely by FAA management. A few of the goals of the D2025 scorecard overlap with DOT Strategic Goals such as the number of severe runway incursions, or accident rates for commercial and general aviation.

During a monthly meeting of the Performance Sub Committee, a performance measure scorecard regarding these goals is discussed. The scorecard rates the status of the targets in green, yellow or red based on the most recent available quantitative data or on the latest qualitative assessment from the manager designated as 'goal owner'. The degree to which management has a direct impact in adjusting results varies from measure to measure. For example the target number for commercial air carrier fatality rate is set at no more than 6.2 fatalities per 100 million persons on board, which equals about maximum of 60 fatalities a year. Luckily, nowadays commercial airplane crashes in U.S. airspace do not occur each year. When they do however this threshold may easily be exceeded. For other targets, management's ability to act and take corrective or preventive actions is easier to imagine such as targets concerning implementation of interventions to mitigate top 5 hazards or to achieve initial operational capacity on the new ERAM system. One of the respondents reflected on this:

'A lot of what we do today does not impact today but may impacts tomorrow. But on this list there are certain things that we can largely control and can put the high level monthly focus on and it does have a direct impact'

Much of the value of the performance subcommittee meeting is credited to the increased visibility of targets within the organization and the signal that goes out from active monitoring by management as explained by one respondent:

'If you have people in an underground parking garage collecting performance data and talking to each other it's much less impactful than if it's the top leadership in the agency sitting down regularly to talk about performance. The folks who are behind each of these measures know that that they are reporting on this monthly and it gets noticed and discussed all the way to the top of the agency'.

There are approximately another 300 measures that are not in NextGen but do appear in the FAA business plan and the business plans of FAA, including the ATO business plan. These

176

measures are mostly tactical and operational in nature and include daily operation metrics for day-to-day management. Some of these measures are tracked every day and add up as annual averages for which thresholds and targets are set in the DOT plans and the FAA Flightplan 2009-2013. The operational, tactical and strategic information is linked to identify trends or developments that may require corrective actions. These kind of metrics concern safety and capacity and largely fall under ATO's responsibility with examples being:

- The number of serious runway incursions (Category A&B)
- Loss of standard separation (the required distance between airplanes in the air)
- The number of operational errors reported and the % that is considered serious
- % availability of systems that support core airports
- Number of daily arrivals and departures (daily airport capacity)
- % on time arrivals in national airspace

At the heart of the data collection for the ATO's performance metrics are a number of vital reporting systems. In table 8.2 four primary reporting systems are summed up:

Table 8.2   Primary reporting systems FAA/ATO reporting systems
            (Source: Mills 2013: 18)

Individual Reporting Systems	
Mandatory Occurrence Reporting (MOR) (1989–Present)	An occurrence involving air traffic services for which the collection of associated safety-related data and conditions is mandatory, including a loss of standard separation on final approach.
Air Traffic Safety Action Program (ATSAP) (2008–Present)	A voluntary reporting program for air traffic control personnel that allows them to enter a qualitative self-assessment of an incident.
Automated Reporting Systems	
Electronic Occurrence Reporting (EOR) (2012–Present)	An alert identified by an automated system, such as the Traffic Analysis and Review Program (TARP) that automatically uploads into the Comprehensive Electronic Data Analysis and Reporting (CEDAR) tool
Terminal Area Reporting Program (TARP) (2012–Present)	An automated software system used to detect and quantitatively measure losses of separation at air traffic terminal facilities.

ATSAP was created in 2008 and was fully implemented by 2010. It allows controllers to voluntarily disclose incidents to an Event Review Committee (ERC). Each Service Area's ERC meets twice a week to review reports in order to determine if corrective actions such as

extra training or procedural changes have to be taken. Under FAA regulations the controller reporting an incident will be provided anonymity and is protected from disciplinary action (Mills 18-21). In fact the safety data reported by employees are gathered by a third party that removes personal identification information prior to sharing it with the FAA (IBM 2011:18)

TARP was implemented in all TRACONS in 2012 and has further been expanding to include *en route* traffic. A group of analysts, known as Risk Analysis Process (RAP) panels, is composed of ATO officials and former commercial pilots who validate the severity of the loss of separation, determine the closing rate of the aircraft, and determine the probability of incident repeatability by examining the state of the system at the time of the incident. RAP panels in each service area meet three to four times a week to examine and validate the highest-risk incidents reported to CEDAR through electronic TARP detection, using a risk and severity matrix (Mills 2013: 24)

The electronic occurrence reports (EORs) generated through TARP and matched with mandatory occurrence reports (MORs) from facility managers and employees provide a rich source of operational performance data. However, due to the confidential nature of the system, ATSAP data are not merged with other sources of FAA data. Both DOT and the Government Accountability Office (GAO) have raised some concern about not integrating ATSAP with other FAA safety data. They fear that the number of serious Operational Errors by controllers may be underestimated and opportunities are lost to pinpoint specific lessons to individual regions and controllers (Mills 2013: 18-22). One major player opposing this has been NATCA, the union of ATControllers which has been a vital partner in the successful implementation of ATSAP.

A steady set of the operational data collected from these systems is discussed daily by ATO senior management as one respondent explained:

'Every morning at 10 am our senior management team gets on a 20 minutes videoconference with our command center in Virginia and we basically run through what happened yesterday in our airspace. We run through all the data daily like delays and how well did we do in managing US airspace given the characteristics of that day like weather. How many delays we had and how well did the fixes that we applied work? That also includes feedback from the airlines. At the end of each day the airlines will tell us whether they thought we did a good job. We have 4 areas of metrics that we track every day: safety metrics, capacity metrics, efficiency metrics and community metrics. Within capacity we for example have daily airport capacity, airspace efficiency, departure delays and arrival delays. '

When asked for which purposes PI was used for within FAA/ATO, respondents indicated all 16 types of use as identified in CH 2.3. Most frequently mentioned were:

- Performance reporting for external accountability
- Setting program priorities
- Analyzing productivity and funding levels

There were followed by:
- Strategically reallocate internal resources
- Monitor cost and performance and contract management
- Allocating internal funds
- Identifying service problems and changing work processes
- Motivate staff to act consistent with organizational goals

The most frequently mentioned types of PI seem to be consistent with the performance management as described to be in place at FAA. It should be mentioned however that due to the scope selected for this cases study, only interviews and questionnaires of people at FAA/ATO headquarters were used.

Finally, the significance of PI for financial allocation use is not entirely evident. Although every dollar is accounted to one of DOT's outcomes, the use of PI for financial analysis seems mostly limited to capital investment decisions and setting adequate staffing levels at facilities. For a large share of especially the Operations appropriation and FAA/ATO's line of work, spending is tied to personnel payments as expressed by a respondent:

'We have the salary budget and the non-salary budget. We dive into productivity metrics for facilities. We also look at metrics that we look at in conjunction with the FAA finance office to efficiently spend our dollars regarding issues such as staffing, overtime, administrative issues'

'We have a really small discretionary part of our Operations budget. Of the 7.5 billion US$ of operations, 85-90% is pay. We can do strategic planning with only about 5-10% of the Operations budget. The rest of it is negotiated pay.... So when the controller's pay goes up, modernization programs are going to be pushed out. As a result things are going to get delayed.'

Although the largest part of the Operations budget may consist of people and pay, the FAA is the only federal agency that has had a relatively longstanding pay-for performance policy towards its personnel. Under a plan called Core Compensation, FAA, the FAA introduced two different programs at the turn of the century. First there is an Organizational Success Increase (OSI) of 1% if the organization meets 90% of 30 indicated OSI goals. This percentage was higher in the past (up to 4,5%) when it included an annual cost-off living upward adjustments but this was suspended due to the crisis. OSI covers most FAA employees. The other initiative in place is the Superior Contribution Increase (SCI) that individual employees

can receive when they exceed expectations and targets from their individual performance work plan. This can amount to an additional 0.8% or 1.6% depending on individual performance and covers about 20-25% of FAA employees.

The incentive from OSI , at least at its current level, is downplayed by most respondents as illustrated by these quotes:

'Because of the budget situation everyone gets that 1%. That 1% isn't going to make me do my work any better'

'Right now we are running we're in our third year without an annual increase. So if we get 90% of our measures successful we get a 1% increase If you take off the tax and deductions we are talking about the difference between a grande or a regular latte at the store'.

Others emphasize that the OSI incentive is more reputation-driven than financial because of the increased focus on OSI objectives at the executive level. On SCI the opinions expressed vary. On one hand it can obviously bring a certain amount of tension among organization members. Therefore it requires that managers formulate clear expectations as well as provide concrete ideas as how to exceed these. According to some it can co-exist or even reinforce an intrinsically motivated and professionally driven workforce as expressed by one respondent:

'My opinion is that employees are motivated by money and by leave because these are tangible. The non-tangible incentives, such as appraisal, are becoming more important as money gets tight. You got to have a passion for what you're doing. If money was the reason our employees would not be in the government to begin with. When you come to work for the civil government you want to make a difference. Having said that, the money aspect really helps and can't be ignored because I think it's a motivator. If you go out to the workforce at the beginning of the year and say to people: nobody is getting a raise because there is no money but keep doing a good job, motivation would suffer. My morale will go down if I work hard and get the same pay increase as my colleague who isn't'

8.2.3 Conclusion on PB implementation

The Operations budget forms the major part of FAA's budget and an even larger part of that of the ATO. This 'going concern' expenditure seems rather immune from performance considerations, at least on an annual basis. Insofar PI plays a role in the budgetary process with FAA's principal, it does so in the budget preparation phase in which OMB seems to be at least as important as DOT as a negotiating partner. All FAA expenses are allocated to DOT's strategic goals up to the detailed level of activities. During budget execution, there are institutionalized performance dialogues between DOT and both FAA and ATO. These center

mainly on strategic goal realization but do not lead to financial reallocation. So although the goal alignment between DOT as a principal and ATO as an agent is quite specific and detailed, this does not seem to translate in direct financial consequences resulting from DOT's management of FAA and ATO.

Within FAA performance metrics do seem to come into play in assessing capital investments as well as staffing of facilities. Although the operations budget in itself has limited discretionary space, FAA has for a long time pioneered a rather comprehensive pay for performance system that ties a small part of personnel pay to organizational and individual performance. The limited discretionary space in a large part of its budget and the need for budget cuts has lead FAA to establishing the Strategy Budget and Performance Committee. The objective of this committee is to prioritize between FAA areas and as such gain an agency wide performance perspective on the possibilities for budget cuts. By cross-cutting the budgetary and organizational stovepipes it is hoped that incremental method to cut budgets, also referred to as salami-slicing, will be avoided.

Table 8.3   Presence of indicators of performance budgeting implementation

Indicators		*Presence in FAA case*
PB 1.1	PB is used by the agency in addition to traditional budgeting	Present to some extent
PB 1.2	PB is used by the principal to control the agency	Present to some extent
PB 2.1	A high degree of de jure PB implementation	Clearly present
PB 2.2	A high degree of de facto PB implementation	Clearly present

The FAA has long been considered an exemplar of how agencies should implement performance management in the government (Mills 2013: 27, Partnership for Public Service/IBM 2011). The emphasis on its data driven safety management is easily noticed when talking to ATO respondents. Monitoring safety data to learn and improve is taken very seriously as illustrated by the daily routine of senior management of going through yesterday's key metrics. The ATO's active data driven performance management approach also extends to its performance dialogue with principals.

8.3   Exploring Neo-Institutionalist Explanations

8.3.1   Explanations from Historical Neo Institutionalism

When investigating path dependency as an explanatory factor, important events with an impact the organization were sought that might explain PB implementation. This could be anything, for example a crisis with a principal or a series of bad press in the media. Moreover it is possible that, if adoption of PB is viewed as an appropriate response to such events, result oriented behavior should be explained by the logic of appropriateness rather than PB's incentives doing their work. To put it differently: organization members are susceptible to unwritten rules and roles that demand compliance with a performance management system, regardless of the incentives of this system itself.

Both the media and the FAA website provide examples a number of important events from recent history that must have helped shape the current FAA as we know it. Deregulation of the airline industry in the late 1970 saw the beginning of an exponential increase in FAA's workload both in terms of airline certification as in the volume of air traffic to be managed. In 1981 an extensively media covered event unfolded as air traffic controllers went on strike after unsuccessful contract negotiations by the PATCO labor union. This resulted in President Reagan firing over 11.000 controllers and handing over part of their activities to military controllers. This left a seriously understaffed agency for quite some time. Eventually it meant the end of the PATCO labor union which was replaced not until 1987 by NATCA, the current air controllers union. The terrorist attacks of September 11th 2001 when four airliners were hijacked and used as flying bombs to attack the Pentagon and the World Trade Center had a profound impact on many in society, not the least on the FAA as an organization. Apart from the extra safety measures that have been in place ever since, it meant that the aspect of aviation security was taken over by the newly formed Transportation Security Administration. This new agency moved from DOT to the Department of Homeland Security in 2003.

Surprisingly these events were hardly mentioned by respondents when asked to name the most significant events that have affected the FAA throughout the years. To prevent answers too close to the respondent's particular duties, respondents were asked to think of answers that they felt any colleague might also give. Ten different events were mentioned by respondents but none by half or even a large part of the respondents. This suggests that no single event is dominant in the collective history of FAA employees. Of the events mentioned, only three were mentioned by more than one respondent:

- Public scrutiny of FAA for badly managed projects and cost overruns around 2002-3
- The creation of ATO as a separate organization
- Failure of the Advanced Automated System in 2000 resulting in enormous delays

Although not mentioned by a majority of respondents the first two events do provide a particularly plausible explanation for much of FAA/ATO's emphasis on performance improvement and performance accountability. From the explanations given, it is clear that the first two events should not be viewed entirely separately. Regarding the first event, several respondents explained that the cost overruns, the badly managed investment projects and FAA's bad press that resulted from it, provided the main driver of FAA's current performance culture:

> *'Around 2002-3 we were getting beat up all the time by the GAO, congress and the Inspector General because they thought we couldn't manage programs. We had acquisition programs that were way over budget and behind schedule....There was a real push in 2002-3 to raise the level of commitment and employ PMI[20] certified program managers for big programs. Any program over $50 million had to have a PMI certified program manager. ...As an agency we were just really tired of being nullified left and right. So we really put an emphasis on stabilizing requirements, building to requirements meeting costs and schedule. And there was a huge a culture shift at that time.'*

> *'Ten years ago we were fat and happy; the budget situation was not as grim as it is today. Our administrators regularly got dragged to the hill and had their head handed to them because we could not deliver capital programs on time and within budget. The numbers were really awful. She cut our 2.5 billion dollar capital program budget with 0.5 billon in one year because of our inefficiency. I asked all executives: What will be the impact on performance). After 72 hours they came back and most of them told me: we are unable to articulate what, if any, impact this will have on performance. Today we can do that in excruciating detail'.*

These stories seem to be confirmed by a 2003 GAO report that indeed is highly critical of FAA:

'*Over the years, systemic management issues, including inadequate management controls and human capital issues, have contributed to the cost overruns, schedule delays, and performance shortfalls that FAA's major ATC projects have consistently experienced. These problems occurred, in large part, because FAA lacked the information technology and financial management systems that would have helped it reliably determine the projects' technical requirements and estimate and control their costs and schedules. In addition, organizational culture issues discouraged collaboration among technical experts and users, and frequent changes in FAA's leadership--seven different Administrators and Acting Administrators in the first 10 years--hampered the modernization efforts*. (GAO-04-227T: Published: Oct 30, 2003. Publicly Released: Oct 30, 2003: *FAA's Modernization Efforts--Past, Present, and Future)*

---

[20] PMI = Project Management Institute (www.pmi.org)

Respondent's suggests that the long-time problems with investment projects to modernize the ATC system, not only explained a shift to performance management, but were one of the reasons for creating the ATO as a new organizational entity in 2003. The ATO was set up with result orientation and performance accountability as core values as one respondent expressed:

> 'The ATO was formed in 2003 from two groups within FAA. The genesis was to run the organization like a business. We need to do better with the dollars that Congress gives us and add value for airlines and passengers. Everyone had to justify every dollar that was spent'.

The suggested causality between the troubled ATC projects at the beginning of the century and the creation of ATO as a performance based organization is confirmed by a 2005 GAO report:

'*The Federal Aviation Administration's (FAA) multibillion-dollar effort to modernize the nation's air traffic control (ATC) system has suffered from cost, schedule, and/or performance shortfalls in its system acquisitions for more than two decades and has been on our list of high risk programs since 1995. FAA's performance-based Air Traffic Organization (ATO) was created in February 2004, in part, to address these legacy challenges.* (GAO-05-331: Published: Jun 10, 2005. Publicly Released: Jul 11, 2005)

An alternative explanation of a favourable attitude towards performance measurement may be found in asymmetry of power such as the dominance of a certain person or organizational unit. When in a powerful position for an extended period such a unit or person can leave a lasting mark on the organization that may include encouraging and promoting the use of performance information.

As argued, claims of poor financial management at the start of the century may have provided an important drive for its current performance culture. Ten years on however, there is no indication that this resulted in an extraordinary powerful position of FAA's Office of Budget or other financial units within FAA. Moreover, the analytic, performance oriented culture that characterizes most of FAA today seems to have started at ATO as component organization before it expanded agency wide (IBM 2011:18). A bottom up development of a performance culture and performance system instead of a top down introduction is not uncommon as this was encountered in other cases studies within this research. That does not mean that leadership of influential managers is to be ignored as an explanatory.

At the level of FAA, the name of Marion Blakey is mentioned as having had a profound impact on establishing performance management at FAA. After taking office in 2002 as FAA's

new Administrator, she made it one of her priorities to use performance metrics to monitor the agency's strategic targets as one respondent explained:

> 'We have been doing the monthly performance meetings for over 10 years. In 2000/2001 we used to have them but the administrator never participated and we only got a handful of executives together. The executives were polite and didn't ask each other hard questions. Then they started getting cancelled. Then a new administrator came in (Marian Blakey) who said I want to run this agency based on metrics and our strategic plan. She was significant in reorienting the agency's approach towards performance management.'

At the level of ATO its influential first Chief Operating Officer (COO) made performance management a clear priority right from the start of ATO in 2003 as he explained in an interview to Avionics Today in 2004 when asked about his personal performance goals:

'For one, I want data-driven decisions by FY2005. That means establish baseline performance goals for safety, service, cost and productivity.
I also want more financial responsibility, which means establishing financial baselines and reports for all ATO components, using cost accounting and labor distribution reporting.
I seek to establish organization excellence--in other words, implement 10 percent of the unfunded portion of the FAA strategic plan through cost savings, reprioritization of projects, and financial initiatives.
And, finally, I'm looking for leadership and human capital management. By Sept. 30 [2004] we will develop Phase 1 of management training--budgets and costs--and ensure completion by applicable personnel'
(Avionics Today, Q&A Russ Chew, September 1st 2004, interview by David Jensen)

It can be argued that the analytical data driven approach to safety within the air traffic control community already provided a strong basis for his agenda. Nonetheless, Mr. Chew's approach did represent a shift in emphasis that reinforced the reliance on a broad set of performance metrics as these quotes from respondents demonstrate.

> 'When I came in over 10 years ago it was always a focus on operational statistics. Not until Russ Chew came in, the efficiency metrics and financial measures were getting bigger. Like his successors, who also came from the airline industry, he wanted us to look from the customer's perspective and spend money accordingly (the airline)'.

> 'When he came in he wanted us to look at the system from a different perspective as well, from the viewpoint of the operation, the efficiency, those things that actually provide direct benefits to the airlines who are our number one customers. Then we started growing, we got more and more metrics that we wanted to track'.

'We had to tie the dollars to everything we did. When the ATO formed it was a 180 degrees shift. Everybody had to re-justify every dollar that you were spending…How is the money that we are spending is going to move us forward. If you can't do that we're going to stop funding that program'.

The emphasis on performance management by ATO's first COO and the FAA's administrator at the time should probably not be viewed separately from the poor reputation that the FAA suffered at the start of the century. As ATO was more or less formed to fix FAA's problem of poor financial management and performance accountability, it is not coincidental that the new leadership promoted a stricter regime in these respects.

There is another way in which asymmetry of power could help explain a strong emphasis on measured performance within an organization. A dominant role of specialists in a certain policy field may explain why quantitative measurement is regarded higher than other, more political, factors. To determine this, respondents were asked to choose between factors that they regard as most influential to FAA/ATO policy in two subsequent questions:

    A) Stakeholder interests   or   B) Political ideas and policy goals
       and
    A) Politics/ public opinion   or   B) Specialist norms

If for both questions answer B is chosen, this would indicate that the policy area is run pretty much by experts and their policy ideas. However, no clear picture emerged when applying this framework to FAA/ATO's line of work. If anything the framework shows that FAA/ATO's field is more stakeholder oriented than it is ideological. Generally the air traffic control system and its metrics are to a large extent quite technical and inaccessible to outsiders. This is reflected by the fact that the FAA workforce is a highly technical one that employs a lot of engineers. The latter point incidentally also explains why the FAA is one of the highest paid federal agencies.

The highly technical nature of the work does however not prevent other parties from having opinions on FAA/ATO line of work. Especially politics is never far away as many citizens are witnessing and experiencing the outcomes of FAA/ATO's work on a regular basis. That may be why, despite the high level of technical expertise required, the FAA has separate OMB people assigned to it to scrutinize the relationship between funding and performance. The experience with political stakeholders overseeing FAA's work has been mixed from the perspective of respondents. On the one hand there is the somewhat unfortunate example of stakeholders misinterpreting the rising number of reported operational errors (see Section 8.2.1). In addition examples were provided of programs and activities that Congress funded but were not in line with the agency's own priorities. In other cases interference by

members of Congress was welcomed as an opportunity to help deliver agency goals. An example of the latter was provided by one respondent:

'Political interference is not necessarily a bad thing. A lot of time the priority also relates to our strategic plan. A lot of people in Alaska got killed in general aviation. The chairman of the appropriations committee was from Alaska urged us to do something about it - congressmen obviously don't like it when their constituents die. We succeeded in bringing fatalities down tremendously and could apply these lessons elsewhere.'

Table 8.4   Presence of indicators historical neo-institutionalism

Indicators		*Presence in FAA case*
H 1.1	Critical juncture in the accountability chain or policy field	Present to some extent
H 1.2	Problem to which PB was seen as a solution	Clearly present
H 2.1	An advocate or champion of PB in a powerful position	Clearly present
H.2.2	A policy field in which specialists are dominant.	Present to some extent

Although not mentioned by a lot of respondents, the events at the start of this century that prompted the establishment of the ATO do provide a likely explanation for the ATO's emphasis on performance management and financial and performance accountability. This emphasis has been reinforced strongly by agency management at that time. Moreover it reported that this emphasis has been spreading to the rest of FAA since. The dominance of specialists does provide some explanation of the dominance of performance measurement and management but at the same time political actors also leave a strong mark on the agency's policies and actions.

8.3.2   Explanations from Sociological Neo Institutionalism

FAA/ATO comes across as an organization that invests heavily in measuring and analyzing performance data and learning from it. Partly this may be explained by the emphasis on safety improvement that seems to characterize the entire aviation industry and possibly also by the engineer mentality that comes with a highly educated technical workforce. Illustrative for this emphasis on performance informed learning are the learning loops for incident reporting, the emphasis on performance monitoring by FAA and ATO management and the active engagement in a performance dialogue with its stakeholders. Respondents could provide numerous examples of performance informed learning, seemingly without much effort. Below is an example of performance monitoring leading to corrective actions as shared by one respondent:

'We are looking for trends based on monthly snapshots. If we get site specific we can track our metrics and keep drilling down to the root of the problem. For example when runway incursions turned orange we had the numbers broken down to specific airports. When we identify them, we send teams out to these airports to see what the problem is. For example North Las Vegas used to have 20, the highest number of incursions nationally although it was only a small airport. The team identified two major factors: there was a flight school and there was a very difficult taxi manoeuvre due to the runway structure. We could put down new markings. At a single intersection 18 of the 20 incursions took place. We marked it with a red circle and named it a hotspot put a description at the side and it was marked on the pilot charts.'

Another example illustrates FAA adopting an agency wide learning approach concerning one of its strategic goals:

'Our general aviation fatalities goal was yellow. The head of our safety organization reported it was yellow and talked about the things they would do to get it to green. We had done scenarios with the administrator about how the performance meeting would go. She turned to the head of air traffic and airports and asked: what can you do to help the head of safety with this goal to get the measure to green. This was very startling because at that time executives who were responsible for a measure felt like this is my measure and you're not going to help me.… Out of that we now run a grants program through our airports office and we were able to actually figure out how to provide grants to airports with the highest number of fatalities to bring them down by doing some improvements in lighting, signing and painting. ATO looked at some of the flight paths that they used, then starting redoing some of those to reduce fatalities. Other parties got engaged sharing accountability'.

Throughout recent years FAA's has seen a development where its data analysis has increasingly been used not only to solve accidents but also to predict them and make improvements in order to prevent them from taking place (IBM 2011:18). A key factor for enabling this pro-active learning cycle has been incident reporting as incidents and errors provide a rich source of risky situations or even near accidents that provide opportunities for learning. This enables ATO to adjust specific procedures or just take basis actions like sending out warning reports to all its personnel regarding a specific risk that occurred somewhere. The introduction of the Air Traffic Safety Action Program from 2008 (ATSAP, see section 8.2.2) was an important development in this respect as it provided strong safeguards to make incident reporting non-punitive for employees. Prior to implementation of ATSAP, reporting errors was much riskier process for an air traffic controller. In the old days, the incidents reported were investigated by the responsible air traffic personnel themselves to see if it was an operational error by the FAA controller or someone else (equipment failure,

pilot error). Subsequently a 25-page report on the incident was completed by the controller and his supervisor and sent to the ATO service center and FAA Headquarters in D.C. (Mills 2013:16).

This process obviously provided a larger barrier to reporting than entering an incident oneself electronically in the knowledge that the data will be treated confidentially like currently in ATSAP. In a 2011 interview FAA Administrator Randy Babbitt commented on the agency's non punitive reporting initiatives:

> '*We also have programs where one can voluntarily report, without fear of reprisal, when something happens. Don't just be silent about it. I'll give you immunity. That's a huge change in our culture, but it's one worth pursuing'.*

The change to non-punitive reporting represented a cultural shift that has been challenging to air traffic controllers. The result of decriminalizing reporting was a strong increase in the number of reported incidents by 2010 (Mills 2013:22-23).

More generally, organizational culture can provide a stimulus for organizational learning or provide barriers that prevent this. When favourable cultural conditions for organizational learning are present, using PI to engage in analysis and debate will be regarded as appropriate behaviour. If the opposite is the case, presenting bad news obtained from performance measurement will not be perceived as a learning opportunity but as threatening to management and to organization members, resulting in suspicion and defensive routines. The conditions for organizational learning were tested among others by letting respondents react to 4 statements:

a. Within FAA/ATO, important issues are being discussed openly and fairly
b. Within FAA/ATO, existing opinions are regularly challenged and discussed
c. If FAA/ATO is being confronted with a problem, a thorough problem analysis takes place prior to taking action
d. If things don't work out in FAA/ATO's policies and execution, lessons are usually learned.

The answers to these statements show that respondents generally feel that FAA/ATO is as a learning organization that learns from mistakes and generally does so after a thorough problem analysis. An obvious exception can be made for tactical issues during operations where an instant response is needed and as is not preceded by a big analysis. From an organizational perspective lessons are often learned from analysis. Being a government organization however means some institutional limitations apply as to how quick lessons can be applied as one respondent indicated:

'We do a lot of lessons learned both form hard data and we play back if there was an incident to see what the causes were and what we could have done differently…. There are some limitations to the how quickly you can implement things and how we are tied to budgets'.

Regarding the cultural aspect of openness, FAA/ATO culture is perceived by respondents to have a reasonably open culture in which results and improvements are discussed easily at all levels. The aspect of reflective openness (challenging existing opinions) scores slightly higher than that of participative openness (discussing fairly and openly) as is illustrated by some of the quotes from respondents:

'In my own personal experience no one is particularly shy about telling what they think of any specific topic'

'From a corporate perspective the worst thing you can do here (with respect to target realization) *is to be green all year and then suddenly flip red. Then someone will ask: what is the matter with your assessment skills'*

'The longer that you're here the more open you are and the freer you are to express your opinion about what is working and what is not. In our meetings everything is said but people that are new to FAA or ATO are more reserved and shy. If I talk to them one on one they will give you some thought and ideas'.

Others note that although that although things are progressing, the desired culture is not there yet entirely. Even if the conditions are there for a safe environment where employees can talk openly about everything some are still hesitant to do so because they feel seniority seems to be somewhat of a requirement to do so.

Organizational culture can also have a beneficial influence on organizational learning if it provides a common understanding across organization members about the usefulness or even necessity of performance data. If this is the case, measurement and reporting these data will be considered an undisputed part of one's job. The effort put into measuring, analyzing and reporting FAA/ATO's metrics indeed seems to be quite undisputed as the elaborate set of measures generally seems to be perceived as relevant by organization members. This even goes for those metrics that are beyond the direct control of FAA itself. Not everyone however seems convinced that using performance measurement is indeed beneficial to FAA/ATO's efficiency and effectiveness. In their answers respondents do confirm that they frequently use PI and are largely familiar with the indicators used for their work. It is important to note that these opinions reflect FAA headquarters were most of this case study was conducted. It seems likely that to the average Air Traffic Controller at an airport may have quite a different perspective as was confirmed by a respondent:

'Air traffic controllers are very involved with their own performance. They move aircraft that is all they are concerned about, not so much our strategic goals. As an agency we make an effort to help individual controllers see how they affect the goals and better realize: this is how I touch the organization and make the FAA successful'.

Finally the professional dedication of employees to FAA/ATO's work should not be overlooked as a factor explaining why PI is regarded a relevant part of the job as expressed by one respondent:

'I have rarely if ever met program managers who didn't fiercely care about their program. There is like a sense of pride, ownership and caring about what they do. It's a sense that it's incredibly important, like a dedication to it. These are ways to document and ensure that they are doing a good job. It is this sense of dedication and commitment, which is more of a cultural thing that is driving excellent performance. I've worked in many places and this is really striking'.

In summary FAA/ATO's culture can be described as one with a strong focus on learning and improvement that has developed even stronger in the last few years as the organization succeeded in utilizing an increasingly amount of PI at its disposal. Performance measurement, analysis and reporting seem to be a normal, integral part of everyone's work and provide a common understanding of the organization's effectiveness. Sharing knowledge that is useful for learning and improvement is being encouraged. Insofar there is any hesitance regarding the learning organization's culture it concerns the perceived ability to speak out, especially for junior organization members. The organization still seems to be progressing in this respect.

Table 8.5   Presence of indicators sociological neo-institutionalism

Indicators		*Presence in FAA case*
S 1.1	Absence learning disabilities	Clearly present
S 1.2	Participative openness	Present to some extent
S 1.3	Reflective openness	Clearly present
S 2.1	Shared view of the meaning of measured performance	Clearly present

8.4   Contextual factors

Just like the U.S. Forest Service case, this case study was done during extremely difficult budgetary circumstances. Not only was FAA operating under continuing resolution[21], later on the sequester also kicked in meaning the implementation of automatic spending cuts for all federal agencies starting with a 5.2% cut in FY13. As a consequence an 11 day furlough (unpaid leave) was implemented for employees resulting in part of the airspace being closed and causing delays at airports. On the other hand the sequestration did challenge some of the existing dynamics when it comes to allocating funding within FAA/ATO as one respondent explained:

> '*We haven't been able to decommission old navaids, old equipment and old towers in the rate we would like to. There are institutional barriers to phasing out the old stuff. Sequestration helps us have a hard look at how to right size our system and make sure we're investing in the future.*'

> '*We are looking at decommissioning old systems, slow down implementation of some of the new systems. We did not want to touch our workforce….Budgets used to be silo's now we allocate more organization wide.*'

Another example of a response by the agency has been the establishment of the Strategy Budget and Performance Committee (see section 8.2.3). The input based budget cuts from Congress has certainly obscured and further complicated the already difficult relationship between funding and performance in FAA/ATO's case. The long term results for strategy and performance of these measures still remains to be seen.

In comparison to LVNL, ATO's smaller sister organization in the Netherlands, some important institutional differences exist as well as differences in the environment in which both organizations operate. First of all, in contracts to FAA/ATO, LVNL has been organizationally separated from the Ministry of Transport for a long time and is directly funded by airline fees.

Schiphol, the main airport of the Netherlands has long been a capacity constrained airport and is situated in a crowded area. This makes capacity and noise exposure more dominant items for LVNL than for ATO. In addition ATO is part of FAA which also runs a large airport improvement grant program, does airplane inspections and manages the schedule times and slots. This arguably allows for a more integrated approach towards improving aviation safety. The confidence that FAA is the nation's, if not the world's leader in aviation safety was expressed by one respondent:

---

[21] Operating under a continuous resolution means that there has not been political agreement on one's budget although the fiscal year has already started. In practice this means that the available budget is a conservative estimate of last year's and restrictions apply concerning new spending.

'Maybe I am arrogant but I think we are not responding to the industry but rather we are driving the industry to be safer through inspection, regulating, and certification. From a safety perspective we collaborate with the aviation industry but we are the drivers.'

None of these institutional and environmental differences has led to difficulties in applying the chosen model for this research or assessing the results of both cases studies.

8.5   Conclusion

The investigation of the indicators for the FS should provide answers to three questions to be answered for each case (see CH 4):

A. *To what extent is this a PB success?*

B *Are the identified conditions present?*

C. *Does context offer a likely explanation?*

Figure 8.6   Numerical scores of the indicators from the interviews and questionnaire results
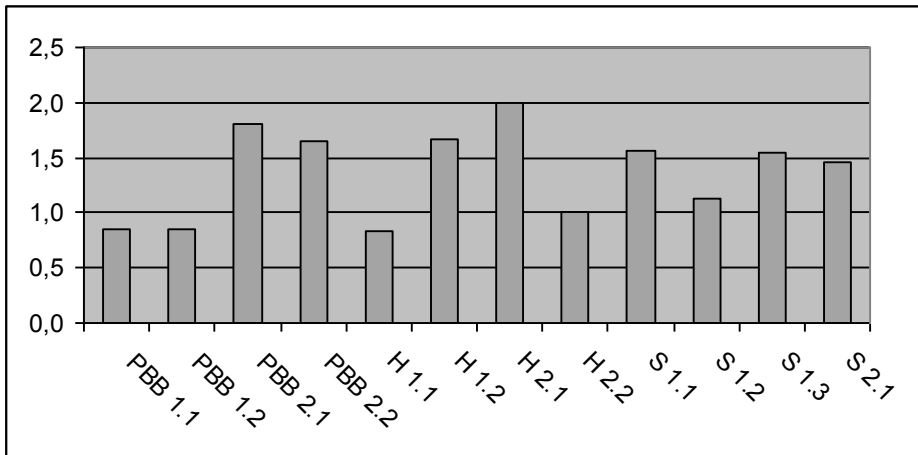
The answer to the first two questions can be given when looking at the numerical scores, obtained from the interviews and expressed in the range of 0-2 and applied to the different indicators. For some of the questions the score was averaged with other results obtained from questionnaires slightly adjusting the final outcome.

FAA's performance budgeting system has some relevance during budget preparation but this relevance should not be overestimated. The reason for this is that there is little discretionary space in the largest part of the budget as it consists of personnel pay. Although expenses are

attributed to DOT's strategic goals, the principals (DOT and OMB) hardly use this information for reallocation funding. It is used for decisions of capital investments. Internally, FAA and ATO use a highly developed system of integration of goals, activities and performance metrics. In this complex and elaborate structure, funding is attributed at the activity level. Under a pay for performance system, organizational and individual performance is both tied to a modest part of employee salaries. Performance metrics from this system are used quite intensively by FAA and ATO during several cyclical performance dialogues that result in corrective and preventive actions.

At the start of this century FAA received fierce criticism by its stakeholders on its apparent inability to properly manage investment projects. As a reaction the President created ATO as a separate performance based organization within FAA. The emphasis of performance management by influential managers of both FAA and ATO at that time should also viewed as a response to this unfortunate episode in FAA history. This event and its aftermath provide a plausible explanation for the appropriateness of performance measurement and management within today's FAA and especially ATO.

The reputation of FAA and ATO as data driven learning organizations is confirmed by this case study. Different learning routines are coupled to performance monitoring and assessment.

The emphasis of learning and improving maybe traced to institutional roots when looking at the history of the entire airline industry although FAA has been pivotal in many of these very developments in the US and abroad. Development towards a more pro-active learning cycle resulting in preventive actions has been enabled by successfully unlocking extra data from incident reporting. FAA's professional and open culture seems to match the profile of a learning organization and provides a likely explanation for the appropriateness of PI use. One aspect that could still be developed further is the notion that everyone, including junior organization members is welcome to debate improvement of effectiveness.

In summary this study revealed evidence that neo institutional explanations indeed help explain the extensive use of PI to learn and improve. Notably important recent events and leadership provide explanations as well as cultural characteristics that largely match those of a learning organization. Contextual factors, notably the budgetary crisis in the U.S. federal government have some impact on the performance budgeting system but provide no likely explanations for PI use at FAA and ATO.

Epilogue
Although not part of the indicators directly tested in this study, James Q. Wilson made the case that a strong professional culture encourage result oriented behaviour as people generally do not need to be told what to do and the flow of information is facilitated by a

high degree common understanding (Wilson 1989:109). Agencies who share a strong sense of mission and a dominant professional culture were even referred to by him as prestigious 'elite agencies' (Wilson 1989:99)

Much of these characteristics seem to apply in particular to FAA culture as encountered in this study. It is therefore surprising to find that FAA ranked relatively low on employee satisfaction among federal agencies. In the 2012 '*Best places to work in the federal government'* research FAA was ranked at place 114 out of 292 federal agencies in overall employee satisfaction with the aspects of leadership, strategic management and performance based rewards receiving particularly low scores (below place 240). In 2009 FAA was even ranked at the near bottom at place 214 out of the 216 agencies surveyed that year. In should be added that being an employer of choice has been an FAA strategic priority since. This seemingly has paid off as FAA today is among the top 40% of federal agencies. Nonetheless this image somehow seems to be somewhat inconsistent with a high performance culture. It should be noted that this case study, a performance culture was observed at FAA HQ and that therefore possible differences with the situation at local stations in this respect went unobserved. Some respondents suspected the pay for performance system was an important factor explaining employee dissatisfaction as the breakdown of today's ranking indeed seems to confirm:

> *'You came here because you had heard we did a good job with using performance measures to achieve results and I think we do. However we also have a kind of a contrary indication in that we are in the bottom of agencies where employees are satisfied with the agency. So it's like we're dissatisfied but despite being dissatisfied we're doing a great job. I personally believe pay for performance is a large part of this because it's so complicated. So people kind of say may say the hell with it I still care about the mission and the programs and I am still doing a good job'.*

Interestingly, this response indicates that FAA employees do a good job in spite of the financial performance incentives provided by its pay for performance system, and not thanks to it. Other responses also indicate that the financial incentives do not affect employee motivation and performance at all:

> *'It is not about the money but about the mission and professional pride. Our salary freeze that we've had for three years now, has not impacted performance negatively'*

In addition to the financial incentives paradox concerning FAA's performance management, this case study also revealed another interesting paradox concerning external performance accountability as briefly explained in section 8.2.1.

Due to the introduction of a non-punitive incident reporting system (ATSAP) and improvements in automatic detection of errors (TARP) the number of incidents reported by FAA/ATO peaked in 2010 and 2012. Ironically, this better data collection initially alarmed external stakeholders—the traveling public and Congress as the number of reported Operational Errors (OE's) was one of the major performance indicator that FAA reported annually to its stakeholders. To them, it seemed that there was a dramatic increase in the number of operational errors and the airspace had become dramatically less safe. In fact, the increased reporting of incidents that had previously been undetected or unreported led to a greater understanding of trends and causal factors, thereby allowing ATO to put in place corrective actions. While this led to a safer air traffic system, it created political problems for the agency (Mills 2013: 22-23).

From the viewpoint of a learning organization these developments were interpreted rather differently by FAA than from the viewpoint of external accountability held by FAA's external stakeholders as this responses illustrates:

> 'The press reported: runway incursions go up by 300. That is not necessarily a bad thing if there is more reporting. In fact it's great that we get extra information that we never got before.'

To add to the irony, in retrospect, despite facing incomplete incident reporting in the old situation, FAA kept missing targets aimed at keeping the number of Operational Errors below a certain threshold. As a response to missing there targets FAA repeatedly made these targets easier to attain. All of this never alarmed FAA's stakeholders. However when the improvement of systems finally disclosed higher reported numbers of incidents that were previously undetected, FAA started receiving a bad press. Throughout the entire period, the number of serious incidents kept diminishing so the extra incidents reported were primarily not severe ones that were nonetheless useful to prevent the serious ones from happening.

Prior to introducing non punitive reporting, controllers' reports indeed tended to be biased in a direction that would make their mistake's severity look less than it may have been. Controllers would be hesitant to report incidents when approaching the annual threshold number set for their unit (Mills 2013: 19-20). Regarding this observation there is an interesting role for the financial incentives form FAA's pay for performance system in this paradox as one respondent explained:

> 'Our learning system has definitely evolved. In previous years we tracked operational data because we wanted to reduce the amount of errors, runway incursions etc. . However these metrics were very public. And so the controllers knew that we had these metric and these also tied into how the performance based systems determined the pay through organizational success increase So we were finding that there was hesitation to report

because they were reporting stuff that could turn around and negatively impact their pay in the end.'

Following discussions between with DOT and OMB a new indicator was chosen to replace the metric of the number of Operational Errors. This new indicator is called the System Risk Event Rate (SRER) and measures the number of serious incidents regarding Loss of Standard Separation (aircraft flying too close to each other) against the total number of such incidents reported.

Data collection FAA case
The primary method of data collection were semi structured interviews. Between March and November 2012, ten persons were interviewed. Two of these worked at the President's Office of Management and Budget and one at the Department of Transportation charged with oversight on performance on key priorities. At the FAA HQ seven persons were interviewed representing management, program analysts, budget and planning staff and ATO management (see Appendix II for more details). Data were primarily processed by qualitative analysis although some quantitative analysis was used for comparing and aggregating results (see section 4.6). These scores were aggregated for each case allowing for a qualitative comparison on an ordinal scale. After finishing each draft case study, the report was reviewed by a key informant from the organization involved. By means of triangulation the results from interviews were compared with the available questionnaire results and findings from analysis of documents sometimes nuancing some of the findings from the interviews.

Sources:
FAA documentation (site, budget ,annual reports administrative factbook June 2012)

GAO-05-331: Published: Jun 10, 2005. Publicly Released: Jul 11, 2005

GAO-04-227T: Published: Oct 30, 2003. Publicly Released: Oct 30, 2003: *FAA's Modernization Efforts--Past, Present, and Future*

IBM Center for the Business of Government 2011. *A Conversation with Randy Babbitt, Administrator of the Federal Aviation Administration*, www.businessofgovernment.org

Jensen, David 2004, *Q&A Russ Chew*, Avionics Today, September 1st 2004, interview by David Jensen

Mills, Russel W. 2013, *Incident Reporting Systems – Lessons from the Federal Aviaition Administration's Air Traffic Control Organization*. Publication of IBM Center for the Business of Government, Improving performance series.

Partnership for Public Service 2012, *Best places to work in the federal government ranking* (survey)

Partnership for public service / IBM Center the Business of Government 2011. *From Data to Decisions – the Power of Analytics*

Wall Street Journal May 2013: *FAA Bonuses and the Sequester*

Wright, Cathy. April 12, 2013, *FAA Bucks Trend in Declining Employee Satisfaction* (http://orginc.com/2013/04/12/faa-bucks-trend-in-declining-employee-satisfaction/)

Zajac, Andrew 2012. Airline Crash Deaths Too Few to Make New Safety Rules Pay (www.bloomberg.com/news/2012-06-25/airline-crash-deaths-too-few-to-make-new-safety-rules-pay.html)

CHAPTER 9 COMPARISON AND ANALYSIS OF CASE STUDIES

The previous chapters 5 through 8 described public sector agencies that use performance budgeting in a way envisaged by New Public Management reforms. The claim to PB success for each case was verified in the first part of the analysis that looked at the agency's internal performance management, the use of PI by principals to control the agency and at the way performance information is used in a budgetary context. Then the analysis of the case studies turned to particular organizational characteristics in each organization's culture and history that may provide a plausible explanation for purposeful use of performance information. In this chapter the overall results from the cases will be presented compared and analysed. First an analysis will take place of the PI use, the neo institutional factors and the contextual factors across all four cases. Subsequently cross comparisons of results will be made between the policy areas (forestry versus air traffic control) and the two government systems (Netherlands versus US). This analysis provides the input for answering the central research question later on in Chapter 10.

9.1   Recapitulation of results from cases

Having conducted all four cases studies the results in terms of the indicators investigated can be compared across cases. To illustrate the variation between case studies and indicators more clearly, the answers from the interviews, which were measured on a three point scale (absent / present to some degree / clearly present), were converted into a 5 way scale in Table 9.1.

Table 9.1   Indicators for presence neo-institutional explanations of purposeful PI use (average from interviews)

	Indicator	*SBB*	*LVNL*	*USFS*	*FAA/ATO*
PB 1.1	PB used in addition to traditional budgeting	+	+	+/+	+/-
PB 1.2	PB is used by the principal to control the agency	-	+	+/+	+/-
PB 2.1	A high degree of de jure PB implementation	+	+	+	+/+
PB 2.2	A high degree of de facto PB implementation	+/+	+/+	+/+	+/+
H 1.1	Critical juncture in the accountability chain or policy field	-	+	+	+
H 1.2	Problem to which PB was seen as a solution	-/-	-	-	+/+
H 2.1	An advocate or champion of PB in a powerful position	-	+/+	-/-	+/+
H 2.2	A policy field in which specialists are dominant.	+/-	+/-	+/-	+/-
S 1.1	Absence learning disabilities	+	+	+	+

S 1.2	Participative openness	+	+/+	+/+	+/-
S 1.3	Reflective openness	+	+	+/-	+
S 2.1	Shared view of the meaning of measured performance	+	+/+	+/+	+

-/- No evidence of presence
- Hardly any evidence of presence
-/+ Present to some extent
+ Present
+/+ Strongly present

By looking at how the cases score on the PB indicators, it can be concluded that the characteristics of these agencies and the relationship with their principals do indeed qualify as cases of successful PB implementation. This goes in particular for the way PI is integrated in budgetary and management cycles and for the way PI is actually used by staff, management and professionals in their work. The relevance of performance information for budgetary purposes and oversight by the principal does not surface equally convincing from all the cases. While results vary per case, the cultural indicators from sociological neo institutionalism seem present quite consistently throughout the cases. In some of the cases, there is clear evidence for presence of indicators from historical neo institutionalism as well. In the next two sections, more specific observations for each of the formulated hypotheses and their underlying indicators are highlighted in the light of the central question posed at the start of the study.

9.2   Degree of PI use and PB success in cases
Although this study took a budgetary view on PI use it was noted that the line between performance budgeting (PB) and performance management (PM) becomes increasingly blurred when investigating the aspect of PI use by agencies for planning and reporting. Not surprisingly, purposeful use of PI is more likely to positively influence effectiveness and efficiency than situations where PI is not used to learn and improve. Even if no direct financial incentives are tied to performance targets and PI is simply included in budgetary documents for presentational purposes, purposeful PI use is likely to have budgetary consequences. These consequences often remain invisible as they may or may not directly show in budget documents. For example if efficiency gains as a result of program learning are offset by higher prices elsewhere, an outsider or maybe even a principal may not notice any change. The same goes for increased customer satisfaction or application of a more effective technique within a giving level of funding. However in the absence of purposeful use of PI for program learning or enlightenment, decreasing service delivery at the same level of funding or increased budget claims would have been likely to have occurred instead. Some claim that due to the lack of success of PB reforms, its definition has been stretched to include broader use of PI or that PB has in fact become a subset of PM (Schick 2014: 3).

Regardless of this debate this study adopts the view that insofar goal setting, performance measurement, analysis and reporting are part of the same formalized cyclical routines as the budgetary cycle, the term PB applies (as may PM at the same time). In doing so this study only focuses on the use of PI that is actively generated, collected and disseminated as part of a formalized performance measurement and reporting system. This can be contrasted with non-routine PI like ad-hoc feedback that is passively received (Kroll 2013: 265).

An important part of the case study analysis was to verify to what extent the selected cases indeed lived up to their reputation of PB successes by demonstrating purposeful PI use in accordance with NPM and principal-agent theory. The result was that in all cases clear evidence of purposeful PI use was found although some variation existed (see Table 9.2)

Table 9.2   Dominant types of PI emerging from case studies. The dark grey cells indicate that over 1/2 of the respondents were familiar with this type of PI use in their organization while the light shaded ones indicate that over 1/3 of them were.

Use of PI for...	SBB	LVNL	USFS	FAA/ATO
Goal alignment or adapting policy assumptions				
a. Performance reporting for external accountability	██	██	██	██
b. Setting program priorities	██	██	██	██
c. Strategically reallocate internal resources	██	██	██	▒▒
e. Understand the impact of external events on performance goals			██	
Better resource management				
e. Deciding on outsourcing decisions	▒▒			
f. Developing and managing contracts				
f. Monitor cost and performance and contract management			▒▒	▒▒
h. Allocate funds to third parties			▒▒	▒▒
Taking corrective actions				
i. Coordinating program efforts with other internal or external organizations			▒▒	
j. Analyzing productivity and funding levels		██	██	██
k. Allocating internal funds	██	▒▒	██	██
l. Identifying service problems and changing work processes		██	██	▒▒
m. Adopting new program approaches following evaluation			██	

Goal orientation and motivating staff

n.	Motivate staff to act consistent with organizational goals				
o.	Setting individual job expectations for staff				
p.	Rewarding staff				

Questionnaire forms filled out by interviewees were used for this analysis except for FAA/ATO where a set of forms was used filled in by a group of ATO lead planners and managers, two of which had been interviewees as well.

As can be concluded from Table 9.2, PI is used most prominently by the investigated organizations in the areas of goal alignment, adapting policy assumptions and taking corrective actions and less for resource management at a more operational level. In addition the case appraisal revealed that the performance dialogue within the agencies was generally much more intense than the one with the principal. Although there is clear evidence of purposeful PI use by all organizations, it can be noticed that SBB does not seem to score as high as its US counterpart. To put the SBB findings somewhat into perspective, the SBB questionnaire (n=85) revealed that over 95% of respondents named at least one type of PI use with all categories being named by in between 6 and 60 respondents. More generally the results regarding PI use from the interviews were confirmed by the available questionnaire results.

To qualify as a 'textbook example' of PB implementation, four indicators were tested regarding the linking of money and results, the performance dialogue, the *de jure* PB system as implemented and its *de facto* use.

Table 9.3 Indicators for PB implementation (average from interviews, converted to 5 point scale)

Score of the investigated agencies on indicators		Average	SBB	LVNL	USFS	FAA/ATO
PB 1.1	PB is used by the agency in addition to traditional budgeting	+	+	+	+/+	+/-
PB 1.2	PB is used by the principal to control the agency	+/-	-	+	+/+	+/-
PB 2.1	A high degree of de jure PB implementation	+	+	+	+	+/+
PB 2.2	A high degree of de facto PB implementation	+/+	+/+	+/+	+/+	+/+
0.0<w<0.4	-/- No evidence of presence					
0.4<w<0.8	- Hardly any evidence of presence					
0.8<w<1.2	-/+ Present to some extent					
1.2<w<1.6	+ Present					
1.6<w<2.0	+/+ Strongly present					

Combining the answers for each case to sub question A: *To what extent is this a PB success?*, based on the evidence from the cases, it can be concluded that the characteristics of these

agencies and the relationship with their principals do indeed largely match the assumptions that were made on successful PB implementation, more precisely:

- PB is used as an integral in the agencies' budgeting processes, most notably for internal allocation, prioritizing, monitoring and external accountability.
- The use of PI by principals to control the agencies varies significantly, is often selective and usually lacks budgetary consequences;
- The organizations have strongly embedded cyclical performance planning and reporting in their budgetary systems (*de jure* implementation)
- The *de facto* use of PI by agencies for the internal performance dialogue and decision making is strongly present in all cases.

A short summary of the PB indicators is provided below. These represent a small part of the overall findings for each case study as described in Chapters 5-8.

Use of PB in addition to traditional budgeting (indicator PB 1.1)
Both forestry agencies (SBB and the USFS) have designed an intricate system of linking funding, activities and outputs and in some cases even outcomes. The SBB framework was introduced in the 1990s and developed further subsequently. The USFS system dates back considerably longer. In fact in the late 1950s Herbert Kaufman in his classical work on public administration refers to the U.S. Forest Service's output based budgeting system as *performance budgeting*, long before this term gained popularity in public sector budgeting (Kaufman 1960:112). The Service's system of linking outputs to funding and labor has been quite stable over decades and in fact Kaufman's description of the FS budget process seems largely accurate of today's budgetary process. The systems of both SBB and the USFS play a clear role in budget preparation and internal allocation as annual work plans and their budgetary consequences are consolidated by headquarters. The monitoring and accountability structure at the USFS is somewhat more intense, as illustrated by the formal accountability provision that requires explanation when targets are not met beyond a 5% margin at all organizational levels. SBB's relies more on annual comprehensive evaluation of a portion of its terrains.

In air traffic control, output estimates like the number of flights do play a role in capacity planning and to some extent PI is used in business cases regarding investments. In budget presentation both financial and PI are presented jointly. Although performance planning does play a role in the activities of both organizations, the bulk of costs of both LVNL and FAA/ATO seem rather immune from performance considerations.

Use of PB by principal to control agency (indicator PB 1.2)
The evidence of actual use of PB by the ministry to control SBB is hardly convincing. Despite frequent contacts, SBB respondents indicated that they seldom got a response from the ministry regarding the PI they reported. This is a development that has deteriorated over

time and is explained by an increasing lack of specialist knowledge at the ministry. The intensity of monitoring in the USFS case appears fragmented as it varies according to policy goal and field of activity. In fact intense monitoring and steering happens quite selectively and concentrates on those activities that are reported externally and receive political attention such as fire, timber, watershed and a limited number of key indicators of the Strategic Plan of the U.S. Department of Agriculture (USDA). Apart from a performance dialogue between the Service's DC headquarters (also referred to as the Washington Office or the WO) and the field, USDA and OMB also monitor the Service albeit with different perspective. In the dialogue with the USDA the emphasis lies primarily on goal alignment and monitoring performance whereas OMB asks questions about effectiveness and efficiency.

Working in a highly technical and specialized industry, performance, results and funding in the case of the air traffic control agencies seems hard to tell apart for a principal, let alone a Congressman or taxpayer. The difficulty of explaining FAA performance to Congress was illustrated in recent years when Congress scrutinized FAA after being alarmed with an increased number of reported operational errors by air traffic controllers. Ironically the increased number of reported errors was the result of improvements FAA made in its incident reporting. This increase of reported errors was viewed by FAA as good news because these incidents previously went undetected and provided valuable opportunities to learn and improve effectiveness (Mills 2013: 27-29). Despite the technical nature of its work, FAA/ATO engages in a critical performance dialogue with OMB, especially during budget preparation and capital investments. The institutionalized performance dialogues between DOT and both FAA and ATO mainly addresses strategic goal realization but do not lead to financial reallocation.

In the Netherlands, the options the government has when it comes to controlling LVNL and its performance and costs, are quite limited given the impact of international and commercial stakeholders. Despite the fact that PI is not directly linked to costs it does play a vital role in the dialogue between the Ministry and LVNL. The ministry, as LVNL's principal, uses PI to reduce information asymmetry and to actively monitor progress towards agreed upon performance targets. Although PI is not used in a direct manner to adjust LVNL's budget upwards or downwards, the ministry has used its power and the available PI, to curb LVNL's tariffs. For LVNL this meant that it was held accountable for the same performance levels in spite of lower incomes due to reduced air traffic, leading to significant cost cutting.

'De jure' PB implementation (indicator PB 2.1)
In both forestry cases the cyclical performance reporting and evaluation is well embedded in the financial planning and control cycle. The USFS system revolves around accomplishment levels that are set for virtually all activities and are tied to fixed funding levels per unit. Accomplishment reporting forms an unavoidable part of the work of almost all employees and forms a stable foundation under each year's budgeting process. The WO does gather

information on almost all of the Service's activities and outputs. In addition tools are used at the regional level that reward over-accomplishment of targets with a larger share of the budget. The SBB system links ecological objectives (target types of nature) and financial consequences is a rather direct way resulting in a more outcome based system. The sophistication of both systems do make it prone to producing a significant bureaucratic burden as information becomes more detailed. Therefore a more pragmatic approach for monitoring is chosen at the headquarters' level by both organizations.

In air traffic control, an abundance of PI is used internally in a systematic manner to improve the organization's effectiveness (notably safety) and, to a lesser extent, for managing financial efficiency. At LVNL, performance is not coupled in a rigid or systematic manner to the budget and the relation between cost and performance is mostly limited to presentation. At FAA, all expenses are allocated to the strategic goals of the Department of Transportation (DOT) up to the detailed level of activities. In addition FAA has for a long time pioneered a rather comprehensive pay for performance system that ties a small part of personnel pay to organizational and individual performance. This does not take away the earlier mentioned fact that the budget of both organizations is rather immune for performance considerations.

'De facto' PB implementation (indicator PB 2.2)
The collection and reporting of PI forms a self-evident part of the job of employees in both forestry organizations. The timber and ecology fields build on a strong quantitative focus which is lacking somewhat in recreation. The cyclical use of PI in the different evaluation cycles in both forestry organizations clearly supports program learning. Without much effort a multitude of examples of these types of program lessons could be named by respondents. This goes for improvements in both efficiency and effectiveness resulting from learning forums, performance evaluation and structural methods of analysis.

In air traffic control the emphasis on its data driven safety management is easily noticed when talking to LVNL or FAA respondents. The data driven quest for safety improvement easily explains why the FAA has long been considered an exemplar of how agencies should implement performance management in government (Mills 2013:27, Partnership for Public Service/IBM 2011). The sophistication of LVNL's performance reporting and analysis system arguably knows no peer in the public sector in the Netherlands.

Although this study focused on the use of PI in the budget cycle, it is clear that the primary process is the dominant driver for cyclical performance reporting and evaluation and not the financial planning and control cycle. In other words: it occurs more as if finance has been linked to output rather than outputs to finance.

9.3 Presence of neo institutional factors in cases

The presence of relevant cultural and historical factors for purposeful PI use as identified in the model was tested in all four cases to provide an answer to sub question B: *Are the identified conditions present?* While results vary per case, the cultural indicators from sociological neo institutionalism (S1&2) seem present quite consistently throughout the cases. In some of the cases, there is clear evidence for presence of indicators from historical neo institutionalism as well.

Explanations from an agency's history were hypothesized into four specific indicators to investigate support for the two H-hypotheses. Some specific observation for each of the formulated assumptions and their underlying indicators are highlighted here for each of the H-hypotheses and corresponding indicators. The underlying observations are described in more detail in Chapters 5-8.

Table 9.4   Indicators for presence historical neo-institutional explanations of purposeful PI use (average from interviews, converted to a 5 point scale)

Score of the investigated agencies on indicators		Average	SBB	LVNL	USFS	FAA/ATO
H 1.1	Critical juncture in the accountability chain or policy field	+/-	-	+	+	+
H 1.2	Problem to which PB was seen as a solution	+/-	-/-	-	-	+/+
H 2.1	An advocate or champion of PB in a powerful position.	+/-	-	+/+	-/-	+/+
H 2.2	A policy field in which specialists are dominant.	+/-	+/-	+/-	+/-	+/-
0.0<w<0.4	-/- No evidence of presence					
0.4<w<0.8	- Hardly any evidence of presence					
0.8<w<1.2	-/+ Present to some extent					
1.2<w<1.6	+ Present					
1.6<w<2.0	+/+ Strongly present					

Hypothesis H1: Path dependency (indicators H1.1 and H1.2)
The FAA/ATO case offers most evidence for the occurrence of critical events as an explanation for purposeful use of PI. As ATO was formed in 2000 as a performance based organization in direct response to the perception of ill financial management, it is no surprise that a strong emphasis was put on performance management, performance accountability and their relation to finance. This event and its aftermath provide a likely explanation for the appropriateness of performance measurement and management within today's FAA and especially ATO. One FAA respondent reminisced:

'Ten years ago we were fat and happy; the budget situation was not as grim as it is today. Our administrators regularly got dragged to the Hill and had their head handed to them because we could not deliver capital programs on time and within budget. The numbers were really awful. She cut our 2.5 billion dollar capital program budget with 0.5 billon in one year because of our inefficiency. I asked all executives: What will be the impact on performance. After 72 hours they came back and most of them told me: we are unable to

articulate what, if any, impact this will have on performance. Today we can do that in excruciating detail'

The other cases also provided examples of relevant critical events but with a more indirect impact on PI use, primarily leading to more emphasis on performance accountability to external stakeholders.

Hypothesis H2: Asymmetry of power (indicators H2.1 and H2.2)
Evidence for this hypothesis was mostly found in the aviation cases, in particular for the presence of an internal advocate or champion of PI use in a powerful position. At LVNL introduction of the dominant performance management system and its further development were helped greatly by leaders in the person of the current CEO. Back in the 1990s, when working as an air traffic controller, he began modelling air traffic control for a thesis. Later on he was responsible for initiating this system organization wide and grew to become LVNL's current CEO. At FAA too, the influence of leaders that emphasized use measured performance data can be traced to certain individuals as this anecdote of a respondent shows:

'We have been doing the monthly performance meetings for over 10 years. In 2000/2001 we used to have them but the administrator never participated and we only got a handful of executives together. The executives were polite and didn't ask each other hard questions. Then they started getting cancelled. Then a new administrator came in who said I want to run this agency based on metrics and our strategic plan. She was significant in reorienting the agency's approach towards performance management.'

At the USFS, like SBB, no clear champion of PB was mentioned in the interviews. When looking further into the organization's history such a champion may have been present after all in the name of the Service's influential first chief forester Gifford Pinchot. He is generally seen as responsible for the design of the Service in almost all of its aspects including the scientific approach to forestry and the system of output funded budgets.

In both forestry and air traffic control, a relative dominant role of specialists was recognizable. The technical nature of work In air traffic control means that recruitment results in a staff with a dominant 'engineers mentality' and a high tolerance for - and understanding of- quantitative measurement. The nature of forestry work may not be as technical as air traffic control in all respects. However the audience for many of the activities (e.g. biodiversity) is rather small if no direct stakeholder interests are at stake. This could also explain a relative dominance of specialists. Although all cases do portray a picture of policy specialists at times being able to go at their business relatively undisturbed by politics and stakeholder interests, politics is never far away to interfere with work and seems to increasingly do so.

Cultural aspects have been regarded as an important factor for PB success and were operationalized into four specific indicators to investigate support for the two S-hypotheses. Some specific observation for each of the formulated assumptions and their underlying indicators are highlighted here for each of the S-hypotheses and corresponding indicators. The underlying observations are described in more detail in Chapters 5-8.

Table 9.5   Indicators for presence sociological neo-institutional explanations of purposeful PI use (average from interviews, converted to a 5 point scale)

Score of the investigated agencies on indicators	Average	SBB	LVNL	USFS	FAA/ATO
S 1.1 Absence learning disabilities	+	+	+	+	+
S 1.2 Participative openness	+	+	+/+	+/+	+/-
S 1.3 Reflective openness	+	+	+	+/-	+
S 2.1 Shared view of the meaning of measured performance	+	+	+/+	+/+	+
0.0<w<0.4 -/- No evidence of presence					
0.4<w<0.8 - Hardly any evidence of presence					
0.8<w<1.2 -/+ Present to some extent					
1.2<w<1.6 + Present					
1.6<w<2.0 +/+ Strongly present					

Hypothesis S1: Cultural appropriateness (indicators S1.1, S1.2 & S1-3)
In all case studies respondents could, without much effort, name examples where systematic use of PI or celebrated learning forums had led to corrections or new insights in recent years. The ability to critically self-assess organizational effectiveness claimed by interviewed respondents was backed up by evidence from documentation and questionnaires. Generally there was wide agreement in interview results on openness as was illustrated by this quote from a U.S. Forest Service employee:

'We have a culture where it seems to be okay to disagree with one another. We have a lot of scientific disciplines working for the same organization. Obviously there is disagreement about what people think that the priority is for a given landscape. It is not seen at disharmonious not to agree on anything. So people feel comfortable to say that they don't agree'

It should be added that the cases also showed that the inclination of employees to engage in an internal dialogue on performance was not entirely free of hierarchy or seniority. Questionnaire results showed slightly lower results on these indicators than the interview results. This may be explained by relative overrepresentation of central management and headquarter staff in the interviews and possibly to some degree of social desirability bias.

Hypothesis S2: Cognitive frames (indicator S2.1)

A shared view of the meaning of measured performance was found to be clearly present in each of the cases, albeit in forestry somewhat more prominent than in aviation. To foresters performance measurement and reporting traditionally is a self-evident part of their job. Backed up by a highly developed professional culture and a strong sense of mission, the relevance of performance measurement is hardly contested as long as its bureaucratic burden stays within reasonable limits. To an air traffic controller the relevance of safety metrics that are directly impacted by his or her actions is more easily acknowledged than other targets within the myriad of analytical metrics that their organizations use for planning, analysis and accountability. A LVNL respondent commented on this:

> 'Working with indicators is widely accepted because the usefulness of some data is clear to everyone. This goes in particular for safety. This is somewhat different for environmental performance measures. If a single plane is in violation of the norms, it is impossible to attain your target for the entire year. This does not motivate air traffic controllers in their daily work'

While evidence from the cases was adequate to answers questions A and B, an explanation of purposeful PI use is not complete without including potential contextual factors.

9.4 Contextual factors

In addition to answering sub questions A and B through the indicators investigated, sub question C: *Does context offer a likely explanation?* was answered for each case in order to explore whether purposeful PI use demonstrated in a case can be explained by other dominant contextual factors than to the neo-institutional ones tested in the model. As context is defined to be everything else not addressed by the model in particular, it would be absurd to claim comprehensiveness with regard to all contextual factors that may also be relevant for explaining the phenomenon investigated. Nonetheless this section will look back on some of the relevant contextual factors that played a role in interpreting the results from the case studies. First however some differences in national context between the US and Netherlands government systems will be addressed to make sure that these aren't ignored when interpreting results for the studies.

National differences

Differences between the US and Netherlands agencies that came up in the case studies concern the national PB system in place, the organizational status of the agencies, the scope of their tasks and some features of the political systems in which they operate.

Agencies in both nations face a requirement to report PI to a principal that has been in place since at least the 1990s. There are however differences in the way performance planning and reporting has a direct budgetary component. As the federal US never formally adopted a

system of program budgeting, the Budget Line Items (BLI's) form the backbone of the financial allocation of both the Forest Service and the FAA. The reliance on BLI's for allocation instead of lump sum financing around policy objectives has long been considered somewhat problematic from the viewpoint of budgetary flexibility and adds significantly to the complexity of linking outputs levels and strategic planning to funding. Indeed linking performance to funding was a lot more complex in the U.S. cases because the structure of budget line items largely disregards performance planning.

Apart from the apparent differences in size and volume, the tasks performed by the Dutch organizations and their US counterparts are highly comparable. With regard to the forestry it should be noted that while fighting forest fires takes up a large part of resources in the US, this activity is hardly existent in the Netherlands, at least in comparison. In addition SBB is the sole government agency tasked with managing forests in the Netherlands whereas the USFS is one of four federal land management agencies. However, as the jurisdiction of these agencies is clearly drawn across geographical lines, no competition or additional complexity from this difference was encountered in the USFS case study.

Organizationally, the two organizations from the Netherlands operated more independently from the ministry under which they resided. Unlike the FAA/ATO which organizationally belongs to the FAA and the Department of Transportation, LVNL has been separated from the Ministry of Transport for a long time and is directly funded by airline fees. This means that the efficiency incentive on LVNL to control its rates to airlines is felt more directly by the organization. Unlike the U.S. Forest Service, SBB also gained independence from the Ministry of Agriculture in the 1990s although SBB remained financially dependent upon the Ministry's annual contribution.

The difference between the political systems became apparent during the case studies in several ways. One noticeable difference was the more complex cascade of policy goals of the U.S. agencies due to the necessity of their performance planning to comply with comprehensive multi-annual departmental strategic performance plans as well as priorities of the President and other political appointees. As agency directors in the U.S. are politically appointed to realize a certain policy agenda, their appointment often directly influences the performance planning and reporting system to some degree. Another feature that distinguished the U.S. cases from the Netherlands is the dominant influence of local constituencies on members of Congress. While preferences of politicians can certainly influence prioritization of policy objectives in the Netherlands, the U.S. cases revealed a more selective political focus on those goals and targets that directly affect local communities. Summarizing these differences it can be argued that the political context in the Netherlands makes for a somewhat less politicized focus on performance issues and a larger influence of the technical perspective of civil servants. Although relevant for

interpreting some of the findings from the cases, none of these international differences led to difficulties in applying the chosen model for this research.

Relevant contextual factors that may influence PI use
In addition to the differences in national context, a few contextual factors were identified that may provide some explanation of purposeful performance use. Some of these are more or less in line with the institutional characteristics of the organizations involved. Others were more characteristic of the environment in which these organizations operate. A contextual factor potentially influencing purposeful PI use in both aviation cases was the observation that the airline industry appears to be a rather cohesive community that shares a clear commitment to safety improvement. This makes the agencies major participants in an industry wide, data driven quest for improving its outputs and outcomes. It can be argued that the institutional and contextual factors are likely to have interacted to a large degree here. For example LVNL's preoccupation with 'just culture' for reporting errors across hierarchy lines is shared with other players in the aviation industry.

A hard to ignore contextual factor affecting all case studies has been the financial crisis as the Great Recession kept unfolding during the period that this research was conducted. This resulted in significant budget cuts hitting the agencies involved. No conclusive pattern of coping with this crisis response emerged from the cases. Generally the crisis seemed to be rightly perceived as a return to input oriented budgeting where the ties between money and results are increasingly cut, or at least temporarily. Some respondents stressed that less funding only slowed the pace of performance planning but not fundamentally affected it otherwise. Others saw cuts as a welcome opportunity to critically assess effectiveness and goal contribution of some of the agency's activities. A particular effect in both U.S. cases was that budget cuts did undermine tools that financially rewarded exceeding a performance target by employee or units. As performance rewards are based on a percentage of total funding, the performance-based increases in fact came down to less of a decrease rendering these tools less effective. Although effects on PI use were observed in the cases, the crisis does not provide a clear comprehensive explanation for an increase or a decrease of PI use. A more elaborate assessment of the responses on austerity with regard to PI use is provided in Chapter 11.

9.5   Comparison results across countries and policy areas
Having compared the individual case studies in order to provide an overall answer to the sub-questions posed, the evidence can also be used to look for similarities and differences at the level of the countries and policy fields studied. If distinct divides between the US and the Netherlands or between forestry and air traffic control do occur this could be informative for explaining successful PB implementation.

Comparing Forestry and Air Traffic Control
Both policy areas are characterized by a complex detailed planning and reporting system that links outputs to funding and in many cases even to outcomes. In addition, the cases show that in both policy areas, important elements of performance reporting have gone largely unchanged for the last 50 years or so. For non-specialists, many aspects of organizational performance are quite inaccessible. In forestry the technical expertise of professionals has been increasingly challenged by public and political interests. In addition forestry has seen a diversification in disciplines from timber to ecology and recreation involving more stakeholders with conflicting goals. Although the different scientific backgrounds present in today's forestry workforce may sometimes be at odds with each other, the strong scientific background of the work and a high degree intrinsic motivation are shared widely by organization members.

Unlike the forestry agencies, both air traffic control organizations saw the use of PI being stimulated by a recent dominant leader who championed performance management. In the FAA and to a lesser extent at LVNL, the emphasis on measured performance can be explained as a response to specific accountability problems. Generally the development and further sophistication of performance management systems was a bottom up process that was initiated from the ranks of professionals.

The cultural indicators (S1&2) do not show any notable divide between both policy areas. Both foresters and air traffic controllers were found to share a strong professional culture which is seen as a favorable condition for PI interpretation and use. This, combined with a large degree of openness and the presence of organizational learning forums, make for an organizational culture that fosters learning and improving using PI. Interestingly, air traffic control has been a leader in the public sector in both countries in developing methods of non-punitive reporting. Having experienced the distorting effect of performance incentives and the fear of legal prosecution on safety reporting, both organizations have proactively sought ways to safeguard maximum availability of data in order to learn and improve.

Comparing the U.S. and the Netherlands
Comparing the agencies in the Netherlands to those in the US can be risky for different reasons, the most obvious ones being the sheer difference in size and the difference in national context of politics and public administration. Although the scores on the indicators do not give rise to clear differences that can be traced back to national differences, they do enable a better understanding on some of the results, notably on those of the PB indicators. Some national contrasts were already noted in Chapter 9.4 across the cases such as the phenomenon of political appointees. A politicized policy emphasis that changes every few years is not known to SBB and LVNL in the way it is to the USFS or the FAA. In addition the alignment between policy goals and funding is more complicated for the U.S. cases because

the agencies do not receive their funding based on programs but based on BLI's. These factors make the alignment of goals with budgets a considerably more complex task.

Another difference is the more decentralized structure of public administration in the Netherlands. This goes in particular for the larger degree of autonomy that the agencies in the Netherlands were formally granted in the 1990s. By relying on a model of decentralized financial stewardship of line ministries and agencies, agencies in the Netherlands are relatively immune to compliance with government wide performance initiatives from the core of government. Therefore their performance dialogue is more internally focused and is limited to the line ministry for external accountability.

A more general appraisal of national contextual differences that surfaced from the cases was addressed in Chapter 9.4. In Chapter 11 further research is suggested on a number of notable similarities between the cases.

CHAPTER 10   CONCLUSION

A familiar sight for many high school students in Finnish schools have become the regular visits of foreign journalists, politicians and education professionals to their classrooms. The Finnish education ministry has staff permanently allocated to arranging school visits by foreign guests. In addition, commercial operators have started to offer trips to the country's schools. The main reason for Finland's popularity as a hotspot for education tourism is the fact that the country tops the international ranking of school performance according to the OECD's Programme for International Student Assessment (PISA). PISA measures worldwide performance of 15 year old students and has become a popular international benchmark for education performance over the last decade. When analyzing Finland's PISA scores however, researchers form Helsinki University mainly identified ingredients for success that bear little relevance to education policy itself such as language, geography and history. Only energetic intervention for struggling pupils was identified as an element that may be replicable in other places (source: The Economist June 26th 2008).

The surge in popularity of performance budgeting by governments and the worldwide replication of reforms that followed somewhat resembles the power of attraction that Finnish secondary education continues to enjoy. While conventional wisdom tells us that what works in one place, may or may not work in another place, this notion seems to have been largely lost to PB reforms (and indeed budget reforms in general). By providing answers to the sub questions and the central question of this research, this chapter will address the impact of institutional and contextual factors on successful adoption of performance budgeting reforms in the cases studies. Far from claiming a universal truth based on a limited number of cases, the results nonetheless bear relevance for assessing the added value of adopting performance budgeting reforms for public sector organizations.

10.1   Back to the central question

In Chapter 1 the international evidence was assessed regarding the mixed results of performance budgeting with regard to transparency (promising) and effective allocation of public spending (disappointing). It was concluded that the potential yield of this reform in terms of more efficiency and effectiveness is most likely to occur at the agency level where policy execution takes place. The general research question posed at the start of this research was therefore:

Are result orientation of a government agency and operational efficiency gains achieved through PB?

This general research question addresses the more tacit and less visible potential results of this reform such intra program re-allocation and result orientation. In order to be able to assess the added value of PB in particular cases, a rather obvious requirement is that cases represent a good practice of having implemented this particular reform. With these considerations taken into account the central research question in Chapter 2 was refined to:

How do underlying cultural and historical factors explain successful PB in government agencies?

This research attempted to answer this question by means of a qualitative comparative case study of cases representing successful PB implementation in a principal-agent relationship. To obtain an answer to the central research question, several sub questions are to be answered. These sub questions are shown in their relevant order in Figure 9.1 and will be followed by brief explanation of the way they were answered:

Figure 10.1   Set of sub questions to be answered

Sub questions derived from central question:

1: What is considered a case of successful PB in a government agencies?
2: What are considered relevant underlying historical and cultural factors?

Questions for each case study:

A. To what extent is this a PB success?

B Are the identified conditions present?

C. Does context offer a likely explanation?
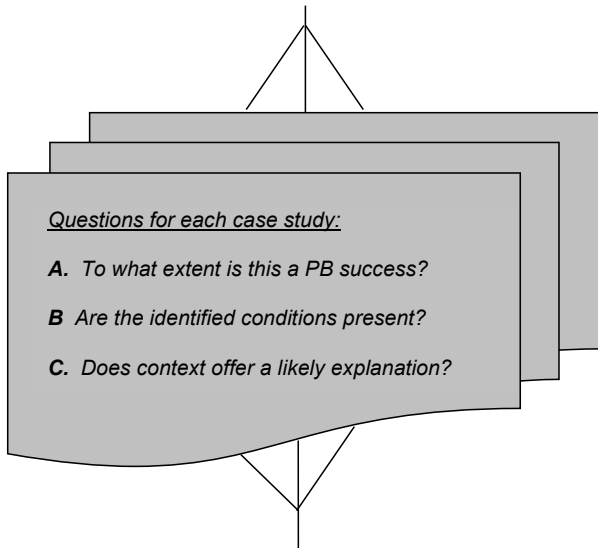
3. Does the presence of these cultural and historical factors offer a convincing alternative explanation for successful PB?

The first two sub questions were addressed in the first chapters:

Sub question 1: *What is considered a case of successful PB in a government agency?*

This question was answered in Chapter 2 with the development a micro model of purposeful PI use by a government agency, both internally and in the relation with its principal. Additionally, four indicators regarding PB implementation were identified in Chapter 4. If PI is used in a purposeful way and PB is implemented in accordance with the indicators chosen, a case qualifies as a case of successful PB implementation.

Sub question 2: *What are considered relevant underlying historical and cultural factors?*
This was answered by identifying alternative explanations for purposeful PI use by government agencies from sociological and historical neo-institutionalism in Chapter 3 in which four types of explanations were identified. In Chapter 4 these were operationalized into eight indicators to be measured in case studies.

After a model was constructed in answer to questions 1 and 2, the sub questions could be answered for each of the case studies:

A. *To what extent is this a PB success?*
B *Are the identified conditions present?*
C. *Does context offer a likely explanation?*

The model constructed directly targeted sub questions A and B through measuring of four and eight indicators respectively. Sub question C on contextual factors explored whether the result orientation demonstrated in a case could be explained by other dominant factors instead of the neo-institutional ones tested. This could obviously not be done with the same level of detail. Nonetheless, answering this question gave some assurance that some obvious contextual explanations aren't ignored and are included in the answer to the final sub question. In Chapters 5-8 these questions were answered for each of the case studies. In Chapter 9 a synthesis across the case studies was provided regarding these answers.

Sub question 3: *Does the presence of these cultural and historical factors offer a convincing alternative explanation for successful PB?*

An affirmative answer to this sub question would indicate that purposeful use of PI can largely be explained by institutional characteristics and is not the result of successful the PB implementation. A negative answer to sub question 3 would imply the opposite, namely that

PB adoption itself should be credited with purposeful PI use rather than the institutional and contextual factors under which this adoption happened.

With due observance of the limitations in terms of external validity that may apply to a qualitative assessment of a small number of cases studies, alternative institutional explanations for purposeful use of PI, were aggregated below into five hypotheses:

If almost no evidence from case studies is found that supports the presence of the defined explanatory indicators, this would imply that purposeful use of PI has little to do with the presence of these indicators and the impact of PB implementation may have been undervalued. This would support the idea that the routines on the collection and dissemination of PI through PB may have modified organizations towards purposeful PI use.

Hypothesis R   Purposeful use of PI is driven by the logic of consequence

Hypothesis R would be in line with the assumptions of the logic of consequence, rational choice institutionalism, (hence the 'R') the principal-agent dilemma and indeed New Public Management. At the very least it would imply that the proper alternative explanatory factors were not identified.

Alternatively, if the cases offer convincing evidence that explains purposeful PI use, the systematic use of PI to improve effectiveness and efficiency has been dependent on a number of unique circumstances and characteristics of the organizations involved. This would be in line with the assumptions of the logic of appropriateness. Four alternative hypotheses were formulated based on the type of evidence to be found in the case studies:

    Hypothesis H1   Purposeful use of PI is explained by path dependency
    Hypothesis H2   Purposeful use of PI is explained by asymmetries of power
    Hypothesis S1   Purposeful use of PI is explained by cultural appropriateness
    Hypothesis S2   Purposeful use of PI is explained by cognitive frames

If there is strong evidence in support of these alternative explanations, the assumed effects of PB reforms should be seriously questioned. Moreover, the causal relationship between successful PB adoption and purposeful PI use may be reversed. Put differently, favorable institutional circumstances for purposeful PI use, are more likely to explain successful PB adoption than the other way around. If this clearly is the case, using PI may have been driven primarily by the logic of appropriateness. This would provide credibility to the idea that, instead of modifying public organizations into PI users, PB reforms codified patterns of PI use that provide legitimacy to organization members.

Hypotheses H1&2 and S1&2 have been proposed as alternative explanations to Hypothesis R. The contradiction between R and the alternative hypotheses mirrors the observation that government agencies tend to be more beholden to their culture and traditions than to budget allocations (Schick 2014: 20). This does not mean that the hypotheses are mutually exclusive as a single one or all may be true. It should be noted that the dichotomy between the two types of logics was used here as a theoretical concept to shed some additional light on untested and often unarticulated assumptions on causality associated with NPM reforms and PB in particular. In fact the hypotheses formulated may overlap to some extent which may not be surprising as, on a theoretical note, historical institutionalism incorporates both a calculus approach and a cultural approach (Hall&Taylor 1996: 5-10).

In addition, the variables used to measure the hypotheses may interact as well to some extent. Therefore it is important to realize that even if the logic of consequence is dismissed as the primary driver of purposeful PI use (rejection of Hypothesis R), this does not entirely discredit the possibility that PB implementation may have has some positive influence on purposeful use of PI at the same time. It can be argued that interaction of variables is to some extent unavoidable when studying complex real life phenomena in social science. Nonetheless whenever interaction of variables was detected or suspected in the cases, this was mentioned explicitly in the case analysis. The cases where this occurred did not alter the conclusions regarding these particular cases. In Figure 10.2 the theoretical model of this research is provided in a simplified matter.

Figure 10.2   Using institutionalism to provide alternative explanations of PI use
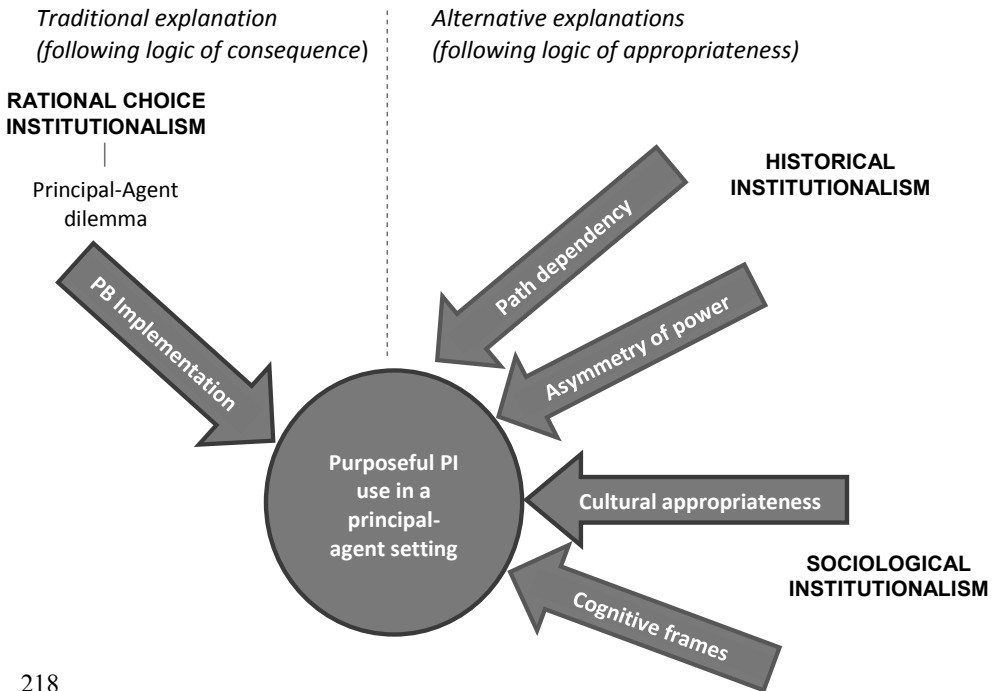
10.2 Result orientation: modification or codification?

Before being able to investigate and explain PB success, this success has to be identified first. In section 9.2 the results from the individual cases on PB indicators were analysed and summarized. In general it was concluded from the findings that the characteristics of these agencies and the relationship with their principals do indeed largely match the assumptions that were made on successful PB implementation. Subsequently, in section 9.3 and 9.4, questions B and C were answered respectively. In short we have concluded that the previous chapter that:

- Cases do generally qualify sufficiently as a PB success (sub question A)
- The neo institutional factors are present in the cases, albeit the sociological ones more consistent than the historical ones (sub question B)
- Contextual factors only offer a limited explanation of PI use in the cases (subquestion C)

Combining the analysis on the four case studies an answer can be given for sub question 3:

Does the presence of these cultural and historical factors offer a convincing alternative explanation for successful PB?

The aggregated results from this study indeed indicate that the institutional characteristics of the agencies examined largely offer a plausible explanation for purposeful use of performance information. More specifically:

- A culture that fosters learning and improving was present in all of the case studies
- Internal systems of performance measurement and reporting are aligned with the professional focus on performance improvement and intrinsic motivation of agency employees at the operational level.
- Institutionalization of a performance based learning system was sometimes reinforced by more or less coincidental events like preferences of a dominant leaders or a crisis regarding financial management or stakeholder accountability.

Combining these findings it seems that purposeful use of PI in these agencies was clearly part of the 'organizational DNA' long before government mandated performance planning and reporting. Looking more precisely at the hypotheses formulated, it becomes clear that evidence supporting assumptions S1 and S2, containing the explanations from sociological institutionalism, are more convincing and represents a more consistent pattern across cases and indicators. The evidence supporting assumptions H1 and H2 vary more between indicators and case studies and are therefore somewhat less evident as an overall explanatory factor. This is illustrated in Figure 9.3.

Figure 9.3 Evidence in support of alternative explanations for purposeful PI use from case studies
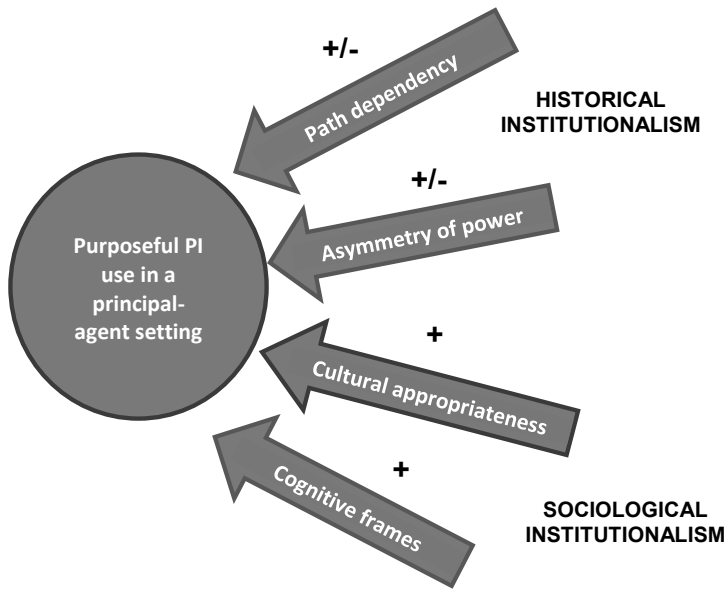
It was noted earlier in Chapter 3 that sociological neo institutionalism can be traced to historical neo institutionalism and is sometimes considered as one of its schools of thought. In accordance with this notion it can be argued that the cultural factors investigated represent a form of historical path dependency in itself as cultures are formed over an extended period of time by factors such as leaders, events and policy characteristics. For the sake of this study these factors were separated and measured as separate indicators.

Now that the neo-institutional explanations for purposeful PI are found to offer a convincing alternative explanation for purposeful PI use, the (refined) central research question can almost be answered:

How do underlying cultural and historical factors explain successful PB in government agencies?

Section 9.2 showed that all four cases studies can indeed be considered good practices of PB implementation and display purposeful PI use. As far as favorable conditions are concerned there is evidence that the factors identified do provide a viable explanation. As stated in section 9.4, with the exception of the safety improvement culture in the aviation industry and possibly the necessity for budget cuts in some cases, no dominant contextual factors were observed that are likely to enhance purposeful PI use (sub question C).

Combining these conclusions it can be affirmed that cases of successful PB implementation (and therefore purposeful PI use) can indeed be explained by favorable conditions that are independent from the adoption of the PB system. This may not surprise many practitioners and is also in line with observations made by earlier by several authors (e.g. Schick 2003: 102, Moynihan& Lavertu 2012: 601). In fact it was Wildavsky who already argued in the late 1960s in response to PPBS (an early PB initiative) that policy analysis and strategic thinking have tendency to thrive on their own and lean heavily on institutional factors such as the availability of capacity. Therefore initiatives to coerce a government wide PB system on government is likely to produce mixed if not very limited results (Wildavsky 1969: 196).

Two notes of caution should be placed regarding the answer to the central question. Firstly, the limited number of qualitative case studies in this research does limit external validity which is why one should be careful when generalizing claims based on these results alone. The representativeness of and variety within the case selection will be addressed in the next section that discusses the lessons to be draw form this research for advancing PB reforms.

Secondly it should emphasized that, due to the complexity of causalities investigated, any comprehensive set of explanation of PI use is likely to interact to some extent. Therefore it is important to realize that even if successful PB implementation may not have been the single driver of purposeful PI use, this does not discredit the possibility that PB implementation may have has some positive influence on purposeful use of PI at the same time. More generally, looking at the case studies, multiple paths to the success of PB seem to exist that might work independently or interact with one another. In one organization, a history of doing performance management might make current approaches more successful. A second organization might lack a strong history, but has a highly engaged leader that makes a difference. A third organization might have both. Therefore, one causal pathway does not necessarily exclude the other. Quantitative analyses generally treats these as independent of one another. A more in depth analysis of individual cases such as this study illustrates the possibility of multiple causal pathways.

10.3   Relevance to performance budgeting ambitions
After an answer was provided for the central research question in the previous section, the potential relevance of this answer to the ambitions of PB and its reform agenda will be the subject of reflection in this session. Let's return therefore to the problem analysis and the general research question posed at the start of this research:

Are result orientation of a government agency and operational efficiency gains achieved through PB?

Now that we know that in the successful PB cases studied, result orientation and operational efficiency did not result from having a PB system as such but rather from other institutional factors, what does this tell us about effectiveness of PB in general? The short answer would be: not much because what goes for these four agencies in two countries necessarily says little about the effectiveness of PB implementation in general. To address the issue of external validity, the potential representativeness of the cases will be discussed later on. Despite the limited representatives of qualitative case studies, these cases have been instructive for gaining insight in the dynamics of PI use for two reasons. Firstly, the case studies provide a clear narrative for the correlation between some of the institutional factors linked to PI use in the public sector by quantitative research. Secondly, despite being viewed as PB good practices, the development and appraisal of PB has at points been at odds with counterintuitive to traditional PB logic.

The issue taken as the focal point in this study was that of modification versus codification (see CH 1.5). In other words, did agencies that do PB well, *modify* their behavior into one that focuses on data driven result orientation by adopting PB? Or did they *codify* their existing behavior patterns into routines just to comply with a formalized PB system? Looking at the evidence from the cases it seems more likely that adoption of PB at some point by the principal, codified existing patterns of purposeful PI use. A similarity across the agencies with regard to their main performance management systems, is the fact that these systems started out relatively long ago and were initiated from within the primary process. Gradually the systems developed into being tools for central planning and reporting by management and staff. Subsequently performance management systems gained an additional role for external accountability to principals and other stakeholders. This order of events somewhat contrasts the traditional idea of PB being introduced by the principal as a tool to control an agent. One may wonder what the added value of having a formalized PB system exactly was for these organizations. More generally, if case analysis of successful PB implementation reveals that PB did not modify these organizations but codified result orientation and extended it to include budgeting, what does this teach us about the potential of this type of budget reform? Before speculating about the answers to these question, it is necessary to discuss the representativeness of the cases studied.

Representativeness of case studies
The limited number of qualitative case studies done for this research does limit external validity which is why one should be careful with generalizing claims based on these results alone. Despite this, the results can be attributed some significance because the cases investigated were selected to represent a small subset of cases where PB seemed to largely function as was intended by NPM budget reformists. Looking at the case selection process there is reason to assume that these kind of examples are not that abundant. When looking deeper into the issue of generalization of findings, the representativeness of the cases studied amidst others in the public sector is obviously an important issue.

One popular classification of agencies is the one described by Wilson (Wilson 1989:158-171) who divided agencies according to the observability of the activities of operators (outputs) and of the results of these activities (outcomes). When looking at the agencies studied, foresters do much of their work in isolation with little opportunity for direct managerial oversight. Therefore the agencies involved rely on ethos and sense of duty to control the organization. As certain results of the forester's work such as putting out fires or harvesting timber can be observed quite easily, the Forest Service could be characterized as a craft organization. Other forestry results however are really hard to observe (e.g. maintaining biodiversity and healthy soil conditions on the long term). In addition the link between these activities and their intended results such as restoring a habitat or improving water quality will only be observable to anyone with specialized knowledge. In fact Wilson referred to the Forest Service as a mixed case of a craft organization as the observability of the service's outcomes varies greatly (Wilson 1989: 167).

As in forestry, the work processes of air traffic controllers are mainly observed by the air controllers themselves when doing their job in physical isolation. The outcomes such as airspace capacity and safety incidents or noise levels are observable and measurable to a great extent. This would qualify air traffic control agencies as craft organizations. On the other hand the processes themselves are (made) highly predictable and highly standardized which is characteristic of a production organization. In addition one may argue that technological advancement has made both the activities of operators as the outcomes increasingly observable. In summary, application of the Wilson framework does not result in clear cut characterization of the investigated agencies. While all agencies seem to qualify primarily as craft organizations, they also have characteristics that resemble coping organizations and production organizations.

Table 10.1   Classification of agencies according to observability of processes and outcomes (Source: Wilson 1989:158-171)

	The activities of operators (outputs) can be observed	**The activities of operators (outputs) cannot be observed**
The results of the activities of operators (outcomes) can be observed	Production organization	Craft organization
The results of the activities of operators (outcomes) cannot be observed	Procedural organization	Coping organization

Referring to Wilson's framework, Pollitt noticed that work related variables have been widely recognized as important to organizational structuring. For example, a 'production' task such as the issuing of driving licenses is standardizable, predictable, and measurable in ways that a 'coping' task such as mental health counseling is not (Pollitt 2006: 5). To what extent the nature of the task of an agency may intuitively be an important predictor of PB success has not been widely investigated. Forestry may have some distinct features that make a forestry agency a relatively suitable one for performance management. Mentioned is the fact that forestry is largely about tangible things as opposed to the application of abstract rules or social influencing social systems. Also the long term character of growing forests and the commercial element of timber sales are seen as favourable in this respect (Pollitt at al 2004: 188/189). Intuitively, the presence of a scientific, data driven tradition in a policy area seems to match quantitative performance management techniques. Both aviation and forestry fall into this category.

Apart from Wilson's framework the selected cases may have lacked some characteristics that intuitively hamper PI use. All agencies studied provide services directly (rather than monitor third parties), and their task is arguably easier to measure than some other functions (e.g. they are not dealing with "wicked problems"). In addition, forestry and air traffic control are not highly contested in a political ideological sense (at least not on a national level) and seem to rely heavily on technical expertise. Although the relative dominance of specialists versus politics in a policy field was assessed by indicator H 2.2, this measurement did not lead to clear conclusive results.

As NPM is popularly viewed as an attempt by the public sector to copy private sector management techniques, PB is often believed to work best in agencies that resemble the classical object of scientific management: the private sector production factory with homogenous and stable products (Jansen & de Waal 2004: 6-7, MinFin 2011: 25,27). However appealing the comparison, this can turn out to be problematic because efficiency in the public sector more often has to compete with various other goals like equity or treatment of politically privileged clients and efficiency may not even be the ultimate goal (Wilson 1989: 175, MinFin 2011: 23).

It can be argued that NPM's key premise of results accountability combined with decentralization (or indeed steering instead of rowing) more easily fits an institutional profile of self-reliance and intrinsic motivation. As a consequence, purposeful PI use and PB may have a 'natural' fit with craft organizations as it traditionally advocates a goal orientation and output and outcome controls instead of input controls. Maybe PB works best in agencies where managers are used to steering instead of rowing, that is, have to leave operators with a great degree of discretionary power and judge them on results afterwards. It is not surprising that successful craft organizations often have a traditional focus on performance

evaluation as a control mechanism, simply because managers and stakeholders can observe outcomes afterwards more easily than processes and outputs of operators.

It may or may not be coincidental that agencies with certain characteristics such as those mentioned surfaced as good PB practices during the case selection in both countries. A more deliberate selection of diversity on characteristics related to policy areas would require a more elaborate analysis beforehand of these characteristics and their predictive value regarding PB success, using one or several frameworks. For now however it is relevant to note that craft organizations appear more likely to share some of the institutional characteristics that explain purposeful PI use. Therefore, when trying to apply the results of this study more widely, this could probably be done most easily for other craft organizations and possibly production and coping organizations. For generalization of the model used in this study, it would be interesting to examine agencies with a successful PB reputation that belong to other categories of Wilson's framework (notably procedural agencies). On the same account it would be interesting to study agencies whose tasks are subjected to perpetual political battles, require little technical expertise or consist of hard to measure results delivered by third parties.

These limitations put aside, the alternative explanations for purposeful PI use as identified and observed in these cases can be instructive with respect to future claims to PB success. As argued previously, the cases investigated were all selected to represent PB successes in the sense that existence of a PB system, purposeful PI use and budgetary consequences all occurred and could be linked to each other in a plausible way. As such these cases represent the vanguard of evidence that PB reformists could use (and indeed did use at times) to demonstrate the effect of PB implementation. If purposeful PI use in these cases can largely be traced to institutional origins of the organizations involved, this would imply that it cannot be easily transferred to other organizations, let alone the entire public sector. In other words anecdotal evidence of purposeful PI use by government agencies should not be mistaken for success of the NPM/PB recipe and does certainly not imply that this recipe can be applied successfully elsewhere. In fact this recipe to modify organizations into more efficient and effective ones may not have been much more than a placebo.

10.4   Reflection on research approach
The approach chosen for this research adds an additional perspective to the knowledge about PI use by government agencies. As asserted in the introduction at the start of this book, a qualitative approach to researching PI use may offer an explanation and better understanding of some of the quantitative work done in this field. In recent years, an increasing number of quantitative studies appeared that linked PI use in the public sector to a diversity of factors adding up to over 30 by 2012 (Kroll 2015: 470-471). Corporate culture, public sector motivation, leadership commitment and discretionary power of managers were already identified as important in this respect (Mahler 1997: 538, Dull 2009a: 273,

Moynihan&Pandey 2010:14, Moynihan& Lavertu 2012: 599-600). Quantitative analysis of data sets can result in important clues on favorable characteristics and circumstances for PI use by public organizations but also has some obvious limitations. Working with a large N-size, inevitably requires simplifying complex variables in order to make them manageable. By relying on abstractions such as pre-determined cultural categories so they can be measured in multiple organizations, no attention is given to specific characteristics and nuances of culture nor to tracing these characteristics to actual PI use. Similarly, another limitation of quantitative work in this area is that it offers little historical perspective on the development of an organization and its PB efforts. Instead, the available cross-sectional analyses and experimental designs have a very short gap between treatment and observed response. For these reasons, qualitative case studies have added value in this field of research.

Especially with complex and intangible factors like organizational culture and leadership, qualitative case study analysis such as this research, are a valuable if not necessary addition to make sense of factors resulting in purposeful PI use. This research should be viewed as a small step to bridge this empirical gap by offering a more specific narrative of the contribution of some of these factors. For example, data availability has been identified by quantitative work as an important factor for performance information use (Moynihan &Pandey 2010: 13, Hammerschmid et al 2013: 7). By using the analysis from this research, it is plausible that data availability itself is a function of organizational history or task.

The theoretical framework
The sociological neo-institutional explanations add a more specific narrative to the widely adopted view nowadays that culture matters, also to purposeful PI use. The historical neo institutional explanations such as the presence of a champion of PB and the occurrence of critical events have hardly, if at all, been the subject of specific exploration in this area[22]. Looking back the choice for neo institutionalism as a searchlight for alternative explanations for purposeful PI use proved quite useful as the contrast between its schools of thought corresponds to a divide in the debate between practitioners. Indeed the division highlighted between rational choice neo institutionalism and its other variants echoes the debate on the limitations of the principal-agent theory and alternative visions such as the stewardship model which was left out of this analysis (Davis et al 1997: 21). Van Thiel 2008: 11). The referral to neo institutionalism may appear a little artificial to some as it is seems to be used primarily to justify the choice for a radically different perspective on the drivers of organizational behavior than the one principal-agent theory offers. Having finished this research does certainly not make the author much of an expert in neo-institutionalism as this, seemingly ever expanding, family of theories seems to comprehend all but the entire

---

[22] Broadnax and Conway (2001: 162/3) do refer to some aspects of agency history and leadership when explaining successful implementation of the 1997 Government's Performance and Results Act (GPRA) by the U.S. Social Security Administration.

universe. The use of hypotheses may be unusual for qualitative research of this type but was seen as useful for clarifying the interpretation of results concerning concepts that by definition seem destined to remain rather vague.

The research design
The choice for triangulation is seen as an absolute necessity with qualitative research such as this. Especially when discussing cultural concepts like reflective or participative openness, the risk of social desirable answers in interviews is a real one as was mildly confirmed when interview results were compared to questionnaire results. The same effect was observed for the level of *de facto* PB implementation from documents when compared to the testimonies from interviewees.

The international perspective chosen is viewed appropriate as the PB reform agenda was and is promoted and adopted worldwide with little regard to national context. The differences between political systems and the organization of government were taken into account when interpreting results. This proved to be a manageable obstacle, partly due to the familiarity of the author with the political systems in both countries from earlier research. In an earlier design the study was to contain six cases studies in three different nations. Although this may have given the results some additional substance it would probably not have meant a significant difference in terms of external validity. Nonetheless it would be interesting to repeat the analysis for a broader variation of international cases.

The indicators used
Had the research been initiated at a more recent date, the base of knowledge from quantitative analysis would have been taken as a more explicit starting point for the model for qualitative analysis. This way more of the results could have been linked more explicitly to the quantitative findings and the underlying data used. It should be noted that many of the factors identified by quantitative analysis were published during or after the empirical part of this research. Nonetheless most of the indicators chosen are believed to shed additional light on selected quantitative research results. The choice of the indicators themselves could in retrospect have included some of the factors that did not became prominent until during the case studies such as professionalism, motivation, physical isolation and longevity of performance reporting tradition. In section 11.2 these factors will be explained more elaborately. The indicator of relative dominance of specialists in a policy field was expected to produce clearer results. One of the limitations of this indicator as operationalized for this research proved to be the fact that some respondents found it hard to apply the corresponding questions to their daily work. This may be explained by the fact that this indicator was based on a framework designed to explain the way issues enter the political agenda and are subsequently decided upon (Conlan et al, 2002: 4). Notwithstanding these somewhat unclear results, the author is convinced that the dominance of specialists in a policy areas versus the influence of things like special interests, political ideology and

media attention contains an important clue for explaining the ability for performance management to develop, grow and mature.

As mentioned earlier, another limitation is interaction of the variables chosen. Related to this issue, and noted earlier in Chapter 4.5, the use of hypotheses for qualitative research concerning a limited number of case studies may be somewhat unusual. This was seen as useful nonetheless for clarifying the interpretation of results concerning concepts that by definition already seem have a tendency remain rather vague. Much of the issue of interacting variables can easily be downplayed by claiming this is unavoidable to some extent when studying real life phenomena in social science. Nonetheless in a few cases a direct connection between different indicators became apparent such as in the FAA/ATO case where indicator H2.1 was likely to interact some with the indicators H1.1&H.1.2. In those cases where a clear interaction was suspected this was explicitly mentioned.

Research method and data collection
Overall, 36 persons were interviewed resulting in over 42 hours of recordings. Interview candidates were selected to represent management, performance & budget staff and operations of each agency as well as staff tasked with oversight on the part of the principal. In addition several experts and senior government officials were consulted on particular issues. Data was primarily processed by qualitative analysis although some quantitative analysis was used for comparing and aggregating results. Each indicator was scored by the author on a qualitative 3-way scale (not present, present to some extent or clearly present) while re listening to the interview recordings. These scores were aggregated for each case allowing for a qualitative comparison on an ordinal scale. After finishing each draft case study, the report was reviewed by a key informant from the organization involved. To limit the risk of biased interpretation by the author even further, it would have been valuable if the interview results would have been scored by several others to see where noticeable differences occurred. Unfortunately there was no opportunity to do this.

By means of triangulation the findings from interviews were compared with questionnaire results and findings from analysis of documents. The availability of questionnaire results did vary significantly form case to case, limiting the ability to use these results for other means than collecting additional opinions as a mean of comparison to the answers given in the interviews. Overall 98 questionnaire results were obtained for 3 out of the 4 case studies. In the fourth case study no permission was granted for issuing a questionnaire. Instead, relevant internal questionnaires were available for consulting on specific results. The reluctance of public organizations to allow this is understandable as employees are sometimes already burdened by questionnaires. Given this reluctance, an alternative approach may have been more effective such as a smaller set of questions than the current 17 and incentives to make participation more attractive.

CHAPTER 11   RECOMMENDATIONS TO PRACTITIONERS AND SUGGESTIONS FOR FURTHER RESEARCH

A rich array of empirical evidence was gathered from the four case qualitative studies. As the evidence was collected with the theoretical model and its indicators in mind, it enabled answering the sub questions and finally the central question. The answers to the questions posed at the start of this research give rise to some lessons and recommendations for those seeking to advance performance based reforms. Additionally, the cases that shared successful PB adoption also shared some other similarities that were not explicitly investigated as part of the model used. Although speculative for this reason, these similarities may provide additional relevant insights on performance management and performance budgeting form a policy point of view and deserve further study. Therefore they are included in this chapter. Finally, as the case-studies were executed while the Grand Recession unfolded, the effect of the fiscal crisis on the working of performance based budgeting and governance systems is also addressed in this chapter.

11.1   Lessons and recommendations for advancing PB reforms

Although purposeful PI use by these agencies may have been more a matter of codification than of modification, depicting the PB system of these organizations as an entirely autonomous development by the agent himself may however not tell the entire story. When applying the evidence from this study to the unrefined research question (see start of this paragraph) it becomes clear that PB's effects on result orientation should be clearly doubted. At the same time, given a result oriented culture and tradition, PB seems to have been helpful with regard to some efficiency gains. At the very least, formalizing the performance dialogue around measured data and linking it to budgetary and management cycles can be credited with several positive effects on efficiency and effectiveness in the organizations investigated. In the cases investigated, PB can be credited by broadly four types of positive effects on effectiveness and efficiency:

1)   *Goal alignment enabled effective execution of agency wide priorities*

First of all organizational effectiveness was helped by aligning the agency's activities to the goals of the principal and those of regional and local units to those of agency headquarters. For all the result orientation and intrinsic motivation that may have been present at the operational level, having a mechanism for systematically linking activities to national policy goals, simply proved to be a necessary precondition for execution of a national policy in the first place. An example of this was for instance the introduction of a national policy for biodiversity in the SBB case.

2) Through an organization-wide PB system, the systemic use of PI already common in one area could be expanded to other, new areas.

Secondly, although the institutional characteristics of these organizations and their policy fields may explain their result oriented behavior for a great deal, it is not self-evident that new organizational goals of a principal will be smoothly adopted throughout the organization incorporated into the daily work of employees. To the contrary, professional autonomy may well prove to be an obstacle in this respect. A particular challenge to public organizations with a strong professional culture is the phenomenon of 'policy stepchildren' which refer to tasks that are not done with the same amount of energy and resources because they are not part of the dominant culture (Wilson 1989: 99). This seems to perfectly reflect the situation as observed in the four cases. For air traffic controllers for example, the environmental and capacity goals appear to be secondary to the professional drive to guarantee air safety. On the same account recreational tasks in both forestry organizations bear little connection to the dominant professions of timber harvesting and ecology. It may therefore not be surprising that performance measurement and results accountability was relatively underdeveloped in these fields. Far from solving the goal ambiguity that faces most public organizations, PB did help integrate new policy goals making use of the structure and result orientation already present in (parts) of the organization.

3) Significant efficiency gains can be realized by allocating capacity according to output data.

Thirdly, having a formalized PB system did add to efficiency by offering a rational basis for capacity planning. Of course one can rightfully argue that sound capacity planning does not require a sophisticated PB system. However the pressures on public organizations to allocate in a more irrational, if not political, manner are sometimes large. As roughly 75% of average agency costs consist of personnel pay, the costs of an internal allocation that is disconnected from production levels can be significant. The effects of rationalizing capacity planning on efficiency was observable in all four cases, with the SBB case in particular offering convincing evidence.

4) Specifying outputs, outcomes and their relationship can smoothen the stakeholder dialogue

Finally the cases show that in terms of stakeholder dialogue, having a common specification of output and performance levels is quite helpful. It is fairly self-evident that a performance dialogue between different parties is helped by a common understanding of, and agreement on performance. Nonetheless the availability of this common basis can be credited to the external accountability aspect of PB adoption. The LVNL case provided a clear example of this.

So summary, in the areas of goal alignment, capacity planning and stakeholder dialogue, PB was intended to play a role and the cases confirm that it can. A formalized performance

dialogue through PB adoption is however unlikely to be a substitute for a lack of purposeful PI use or a low level of non-routine use of PI altogether. Put differently, the good news is that PB can sort effects that benefit effectiveness and efficiency of public organizations. The bad news is that this isn't likely to benefit just any public sector organization. Moreover it seems that, as argued before by others, those organizations that need it least (e.g. have a professional result oriented culture) are most likely to benefit and those that need it most (lack such a culture) are least likely to benefit from PB. The important notions to take away from the results in terms of implications for policymakers and practitioners in the areas of PB and PM in government are two additional points of caution when assessing good practices summarized below:

5) *A PB good practice may not tell teach us much about successful reform implementation*

A claim of a PB good practice should be met with a healthy degree of suspicion and should not go without an assessment of autonomous favorable preconditions in the areas of organizational culture and institutional history

6) *Good practices of performance budgeting are unlikely to be transferrable to others*

Formalizing PI and linking it to funding can have beneficial effects on goal alignment (and therefore effectiveness), efficiency and stakeholder dialogue in some organizations. However these effects may not be likely be reproduced in organizations lacking a data driven performance oriented background and an organizational culture that supports learning.

A call for more caution towards the assumed universal potential of PB good practices and should not be misunderstood as a total disregard of the potential benefits of PB. In the author's opinion academic studies often fail to avoid the pitfall of presenting PB evidence in an advocacy manner. Some of these studies cynically portray ill-advised politicians and civil servants who willingly follow the naïve yet lucrative ideas of consultants or fellow bureaucrats that have been proven wrong over and over again. 'Believers' on the other hand employ scientific methods to lend credibility to their claim that PB as a method has the capacity to change the political process of allocating funds and things as obstinate as organizational culture. It is this author's opinion that the debate on the usefulness of methods to advance evidence based public management would be helped if practitioners would adopt some more skepticism and academics would show some less cynicism.

Implications for the potential of government wide performance based reforms

Performance oriented reforms such as GPRA, PART and GPRAMA in the U.S. system and VBTB, agentification and the introduction of quasi contracts with ministries in the Netherlands all share the notion that mandating performance planning, measurement and reporting would make government more effective and efficient. Additionally NPM assumes that by substituting input controls for output and outcomes, as the PB recipe prescribes,

managers gain discretionary power to achieve the results they are held accountable for. Perhaps surprisingly, some have noted that the assumption that the NPM-ideal type agency model[23] enhances performance of public sector organizations remains largely untested as the claim of an increase in quality and efficiency of service delivery has been often assumed but seldom well documented (Verhoest et al 2012:4). With regard to using PB for agency management this is even more surprising since PB has been promoted as a solution to smooth if not solve information asymmetry, the central problem in agency theory[24].

Where should the evidence from this research be placed amidst other assessment of the impact of PB reforms regarding agencies in a principal agent setting? First there is some evidence from quantitative research that PI use is positively linked to a number of organizational features that are clearly reflective of the NPM/PB ideal. Among these are the requirement to report to a parent ministry (Laegreid, Roness& Rubecksen 2008: 52), decision flexibility from managerial autonomy (Moynihan&Landuyt 2009: 1101) and availability of performance data (Moynihan&Pandey 2010: 13). A recent survey among continental European governments even suggests that PI availability was the most relevant variable for manager's use of PI out of eight organizational factors tested (Hammerschmid et al 2013: 7). Therefore it can well be argued that a consistent government wide PB reform can help strengthen these characteristics and may positively influence PI use.

On the other hand, the fact that PI use correlates with availability does obviously not imply that availability will therefore lead to use. Indeed there is ample evidence that production of PI is not the same as using PI, let alone using PI in a purposeful way as envisaged by reformers. Most practitioners will easily confirm that much of the PI that is produced remains unused, which was phrased accurately by Moynihan:

'Governments have never been so awash in performance data, mostly because they are required to collect and report it. The wealth of performance data contrasts with the poverty of the theoretical and empirical justifications for performance-reporting requirements. We have poor theories of PI use, largely informed by a combination of common sense, some deeply felt assumptions of how government should operate, and a handful of success stories'. (Moynihan 2008:5).

---

[23] This NPM-ideal type agency model is considered to be a model based on performance contracting between a principal (e.g. a ministry or agency headquarters) and a public agent (agency or agency unit). The principal exercises performance control and uses PI for budgetary purposes while the agent is managed in a performance driven manner.

[24] Principal-agent theory is concerned with ways to get an agent to behave in the interest of a principal. The central dilemma in principal agent theory is information asymmetry that occurs when the agent has an information advantage over his principal. The agent thus has the possibility to serve its own interests at the expense of those of the principal. According to public choice theorists the agent can be expected to behave this way, leading agencies to behave inefficiently.

In fact not everyone agrees that formally mandating PI production is at itself as beneficial as one may intuitively think. Some argue that central government initiatives may even absorb the scarce existing capacity from data analysis to compliance with reporting requirements (Posner&Mahler 2012: 3). This way it is even imaginable that the emphasis of PI use may shift form purposeful use to compliance. The latest federal US initiative, the 2010 GPRA Modernization Act (GPRAMA), introduced a series of routines to encourage PI use by agencies. For example GPRAMA mandates agencies to engage in quarterly data driven reviews of performance information for assessing priority objectives (Moynihan&Kroll 2015: 9), not unlike the routines that existed in the organizations investigated in this study. In a 2014 evaluation of GPRAMA, GAO claimed that PI use by agency managers was down compared to their 2007 survey (GAO 2014: 9). Interestingly Moynihan and Kroll, using the same data-set, demonstrate that while overall PI use may be down from 2007, GPRAMA did positively influence PI use and that purposeful use in fact was affected positively (Moynihan&Kroll 2015: 24-26). So with regard to the impact of adopting mandatory top down PB systems and introducing routines for PI use, conflicting views exist that can both be backed up by evidence.

When looking at the conclusions of this research one may wonder why central governments seem to be investing heavily in performance routines to *nurture* purposeful PI use if PI is largely embedded in the *nature* of public organizations. Taking into account that one should be modest with generalizing from a few case studies as well as the availability of evidence that links PI use to PI production and reporting, a few recommendations do come to mind:

7) *The goals and indicators of a performance system should reflect the professional goals and ambitions of agency members*

To harvest fruits of purposeful PI use by government agencies, any attempt to formalize PI use in a cyclical system should remain close to the sense of professionalism and the measures of success of organization members at the operating level. It does not seem like a coincidence that the most successful and long lasting performance systems encountered in the cases were created bottom up by agency professionals from the work floor and not coerced upon an agency or unit in a top down matter.

8) *Simultaneously address aspects of organization culture associated with learning culture*

While cultural variables such as reflective and participative openness were major explanatory factors for PB success in these cases, it seems not very useful to just assume that these will be there in any given agency. Although organizational culture cannot be organized or altered easily on the short term, it may be worthwhile to investigate what existing cultural elements may be useful for, or may provide an obstacle to purposeful PI use. In addition it is instructive to closely observe the response of managers to introducing

elements of a performance culture such as learning forums or challenging employees to discuss performance with their superiors. This may give away useful early signs for the likeliness of PB success in a particular institutional setting.

11.2   Discussion and suggestions for further research

As this research was intended to be partly exploratory and resulted in rich empiric material, a number of striking similarities between the cases surfaced and are worth mentioning. These were not tested directly in the model used and may be considered 'by-catch' from this research. Nonetheless they may provide viable explanations for purposeful PI use and therefore may hint to additional relevant explanatory factors as they were shared by all four organizations. These explanatory variables may partly be intertwined and are identified as: long term reporting tradition, physical isolation of operators and intrinsic motivation. As these three factors have not been the subject of systematic analysis one can only speculate about their true value as an explanatory factor for purposeful PI use in these and other cases. In future research these factors do deserve further exploration and testing. The section will end with an assessment of the impact of the crisis on PI use in the investigated cases.

Long term reporting tradition

It is notable that in forestry and air traffic control, output measurement and performance reporting stems from a long standing tradition. Although the learning system has evolved significantly over time in terms of sophistication, respondents indicated that the air traffic control profession has known a long standing tradition of performance reporting and self-learning. In forestry quantitative planning methods in the lumber-industry and terrain management have been around long before the modern concept of performance management. In ecology, the presence and abundance of indicator species are traditionally used as an indication of the well-being of a larger group of species. The continuity of output measurement was confirmed by respondents of both SBB and the USFS. Regarding the latter Herbert Kaufman, in his classic study of the U.S. Forest Service in the late 1950s, already noticed the existence of countable work output, quality standards and standardized lists of specific jobs as well as the needed amount of time to perform them (Kaufman 1960: 115). Moreover, the description of the USFS budget process in the late 1950s by Kaufman seems to be largely accurate of today's budgetary process.

Physical isolation of operators

In an assessment of the PI use by Norwegian state agencies, Laegreid et al expected to find a lower degree of performance steering in agencies with a territorial component (Laegreid et al 2008: 52). The reasoning was that such an agency operated more autonomous as it was more heavily embedded in regional and local networks. As a result its principals were expected to exercise less control over its activities. This relationship turned out to be inversed in his study although the relationship was discredited as weak when controlled for

234

other variables. The reasoning may however be valid when applied to the relationship between an agency's administrative headquarters and isolated territorial operating units. In fact Wilson (1989) has characterized such an agency as a craft organization where outputs cannot be observed by distant managers. It may only be logical that performance reporting has a long standing tradition is such an agency to solve this very problem. Adding of a top down component of performance reporting to a line ministry may therefore result in a relatively smooth fit. This has been observed in all four cases.

Some have argued that due to the nature of the tasks performed by a Forest Service, 'tendencies toward fragmentation' such as geographic dispersion and varying local conditions pose extra challenges to attaining organizational unity, compliance and conformity (Kaufman 1960: 86-87, Tipple&Wellman 1991: 422). Admitting these factors are physically far greater for the U.S. Forest Service than for SBB, it may equally serve as an explanation of organizational culture of both organizations. Out of necessity, management traditionally relied heavily on the autonomy and discretion of foresters to manage their own terrains. In Kaufman's assessment, the U.S. Forest Service has traditionally relied heavily on training and indoctrination a tools of integration to counter the apparent tendencies for organizational fragmentation. This resulted in a strong cultural and professional identity and a strong sense of mission of the FS.

Not unlike the forester, the air traffic controller does his or her work with a great degree of autonomy from headquarters and has to rely largely on self-motivation. In addition to the physical isolation during work, the professional autonomy of an air traffic controller is further strengthened by extensive in-house education and a comparatively direct and dramatic impact of failure on the job.

Intrinsic motivation
The intrinsic motivation and professional culture is an apparent feature that surfaces from the case-studies. It has been noted by some that PI use and performance positively correlate with intrinsic motivation of employees for the organization's tasks or public service motivation (PSM) as it is often fashionably referred to these days (Perry&Hondeghem 2008:8, Moynihan & Pandey 2010:11). One might argue that a passion for monitoring the results of one's work perfectly fits any performance management system. It does not seem unlikely that an intrinsically motivated person would be interested in the results of policy measurement and opportunities for policy improvement.

In an organization where regional management and headquarters traditionally had to rely on performance reporting by isolated units, it is not surprising that devoted and self-reliant (if not obstinate) individuals fare best. Intrinsic motivation and a clear and stable sense of purpose have had to make up for the absence of an intense system of real-time monitoring by management because this was never a realistic option. If one is looking to define result

orientation within public sector organizations, intrinsic motivation, self-reliance and a clear sense of purpose may be key elements. One might even argue that NPM's key premise of results accountability combined with decentralization (or steering instead of rowing if you will) more easily fits an institutional profile of self-reliance and intrinsic motivation.

11.3   Impact of the crisis

The influence of the fiscal crisis was mentioned in Chapter 9 as a contextual factor and deserves some explicit consideration. The relationship between the crisis and PI use has only recently been the subject of analysis and requires additional research. The Great Recession that kept unfolding during the period that this research took please resulted in significant budget cuts hitting all the four agencies examined. No conclusive pattern of coping with this crisis emerged from the cases. In the final part of this section agency responses that were encountered in the case studies are categorized into four reactions:

I.   View the crisis as an opportunity resulting in performance informed budget cuts
II.   Rely on performance informed coping strategies to deal with budgetary uncertainty
III. Financial performance incentives become meaningless and loose appeal
IV.   Performance contracts with principals lose their value as a basis for budgeting and are replaced by reliance on professionalism in the field

It is important to stress that the effects of the crisis on PI use by agencies does requires further study and the classification below should therefore be regarded as a first attempt to make sense of the observations. In addition several types of responses could be observed within a single case study.

I.   An opportunity for performance informed budget cuts

Like the U.S. Forest Service the FAA has been operating under continuing resolution[25] through most of recent fiscal years. In addition, from 2013 the sequester[26] also kicked in for the USFS and the FAA, meaning the implementation of automatic spending cuts for all discretionary spending starting with a 5.2% cut in FY13. As a consequence an 11 day furlough (unpaid leave) was implemented for FAA employees resulting in part of the airspace being closed and causing delays at airports. The input based budget cuts from Congress have

---

[25] Operating under a continuous resolution means that there has not been political agreement on one's budget although the fiscal year has already started. In practice this means that the available budget is a conservative estimate of last year's and restrictions apply concerning new spending.

[26] The Budget Control Act of 2011 established enforcement mechanisms to reduce federal budget deficits by at least $2.1 trillion over 10 years. The act mandated automatic spending cuts for most federal government departments and agencies, if the Congress failed to enact balanced deficit reduction legislation. These budget reductions, known as sequestration, began on March 1, 2013, and are slated to last 10 years. While exempting most mandatory programs such as Social Security, Medicaid, federal pensions, and veterans benefits from cuts, the 2013 sequester reduced most discretionary budget accounts by approximately five percent, or $85 billion for the government as a whole (source CBO)

certainly obscured and further complicated the already difficult relationship between funding and performance in FAA/ATO's case and the long term results for strategy and performance of these measures still remains to be seen. On the other hand respondents also indicated that the sequestration did challenge some of the existing dynamics when it comes to internally allocating funding enabling to settle long lasting debates on for example decommissioning old outdated equipment.

With regard to FAA's strategy to implement these budget cuts, respondents indicated that there was a sense in FAA that it would be better to move away from salami-slicing techniques and across-the-board cuts. Zero based budgets were not felt to be a serious option because this would be too far from political reality. The limited discretionary space in a large part of its budget and the need for budget cuts has lead FAA to establishing the Strategy Budget and Performance Committee in 2012. The objective of this committee was to prioritize between FAA areas and as such gain an agency wide performance perspective on the possibilities for budget cuts. By cross-cutting the budgetary and organizational stovepipes it was hoped that incremental method to cut budgets, also referred to as salami-slicing, will be avoided.

At LVNL budget cuts have also been implemented in air traffic control albeit with a more active role of LVNL's principal, the Ministry of Infrastructure. Although performance reporting really does seem to fill (or at least ease) an information gap in technical expertise between principal and agent, the budgetary relevance of most PI reported to the ministry seems to be rather limited on a first glance. This results from the fact that virtually no public funds are transferred directly to LVNL. If one takes a second look however, the ministry, working closely with other stakeholders, has made a successful effort to curb spending of LVNL in support of the policy goals of the government. In response to the looming financial crisis, LVNL's projected incomes had to be adjusted downwards dramatically during 2009 due to lower traffic volumes. Out of concern for the aviation sector in the Netherlands and the competitive position of Schiphol among surrounding international airports in Europe, the Minister decided to restrict the tariff rate development for a number of years. This effectively prohibited LVNL from translating operating deficits resulting from lower operating volumes of air traffic, into higher tariff levels. To meet the reduced budget, LVNL set in motion an unprecedented cost cutting operation that resulted in the reduction of 128 FTE, over 12% of staff levels. The reduction in personnel did not affect the number of air traffic controllers as it was filled in entirely by the other LVNL units. It can rightly be claimed that the ministry did not bring about this efficiency operation at LVNL all by itself as she was helped by other powerful stakeholders. On the other hand the Ministry was quick to point out that other member states saw the costs of their air traffic service provider gone up in spite of the crisis.

II.   Coping strategies for budgetary uncertainty

A complicating factor for the performance budgeting system to function for U.S. federal agencies is the fact that agencies usually have to wait well into the fiscal year to know what their definite allocation of budgets and targets will be. At the regional or local level this delay is often even longer as agency headquarters start internal allocation only after a budget deal in Congress is finally reached. The U.S. Forest Service has learned to cope with budgetary uncertainty using an ingenious system that seems to have been in use throughout the regions for a long time at the National Forest level). This system is based on competitively prioritizing combinations of targets and funding at different levels of funding and seems to prepare the agency well for the additional insecurity brought by the crisis (see Section 7.2.2 for details). Surprisingly this approach resembles quite closely an innovative technique recently proposed in an influential paper by Mario Marcel on creating fiscal space that has been extensively discussed in the OECD and World Bank in recent years (Marcel 2013: 32-33)

At FAA the Operations budget (the largest source of total FAA and ATO funding) was frozen by Congress for a number of years prior to the sequester. There was no scaling down of ambitions in terms of outputs or policy outcomes. As negotiated pay for air traffic controllers went up, modernization programs were starting to get pushed out. To deal with this growing asymmetry between ambitions and funding, in 2012 the Operations Review Board was initiated. This exercise illustrates FAA's systematic approach at assessing and reprioritizing budget-levels as it prioritized the operations budget into 3 layers:

- Hard operations which refers to what is needed to keep operations running on an everyday bases (lights, electricity, fuels, leases) and is funded 100%.
- Operations programs that include running and planned modernization and life extension programs
- Operations support: basically consists of everything else, mostly Headquarters and contractors.

Perhaps not surprisingly most cuts were decided in the third category.

III.   Financial performance incentives lose their appeal

Within the U.S. Forest Service, Pacific Northwest Region 6, the region studied in the USFS case, has a reputation of taking performance budgeting quite seriously. A reason for this is the development of a local performance budgeting process in which past performance plays a role in the distribution of next year's budget among the region's forests. According to this system, which was implemented in 2007, a unit can slowly increase its share of the total budget by repeatedly accomplishing more than their targets. Oppositely, continuous underperformance is punished by a decreasing share of the budget. The system is well balanced and can have a serious impact on forest management. Due to decreasing budgets however, the incentive that results from this tool was reported to be felt less clearly by

respondents. Of course one can argue that, rationally speaking, receiving less of a budget cut than your colleagues could still stimulate people to perform extra. This was however not perceived as such by the employees that were interviewed.

At FAA comparable dynamics were reported. The FAA is the only federal agency that has had a relatively longstanding pay-for performance policy towards its personnel. Under a plan called Core Compensation the FAA introduced two different programs at the turn of the century. The most important one was an Organizational Success Increase (OSI) of 1% for all employees if the organization meets 90% of 30 indicated OSI goals. The financial incentive from OSI was reduced from up to 4.5% to just 1% after its cost of living adjustment component was scrapped due to budget cuts. The impact of the incentive at its current level was downplayed by respondents as materially insignificant and unfit to motivate them into doing their work any different. Others emphasized that the OSI incentive is more reputation-driven than financial because of the increased focus on OSI objectives at the executive level.

IV.   Performance contracts lose their value for budgeting

Arguably the most fundamental implication of the crisis on the use of PI for budgeting is that budget cuts are by nature input oriented and tend to ignore the intricate system that intertwines funding, activities, outputs and expected outcomes. In a sense, policy fields and agencies that rely most on rationalizing their budget this way (e.g. the ones selected for this study) are most vulnerable to political leaders and principals that have less to spend, but do not willingly accept rational implication like reductions in their 'production orders'. From the case studies this became apparent in several ways.

The impact of the fiscal crisis has landed particularly harsh on SBB with intended budget cuts of over 25%, understandingly leading to budgetary insecurity. This insecurity was reinforced by the move to decentralize most SBB tasks from the Ministry to the 12 provinces leading to an increase of the number of primary principals to deal with. This has had a negative impact on the interest for output reporting from the side of the Ministry. The system of output controls and associated price levels was agreed with the Ministry in 2007 after a few of years of conflict during which the system was audited twice to check the validity of cost prices. Despite this, the ministry's annual contribution only covered between 50 and 60% of these cost prices since. Ironically, upon completion of the new system of output steering, the major budget cuts and decentralization were announced, undermining full implementation of this new methodology.

As a compromise to deal with underfunding of desired outputs, the ministry and SBB annually specified work that SBB would attempt to deliver but would not will not be held accountable for. This output was included in annexes to the annual performance contract. This implies that PB methodology is taken seriously by the ministry albeit in a way rather different or even inverse from the image of the principal controlling the agent. In response

to underfunding the agreed upon targets, SBB managers in the districts and terrains use their professional expertise to set their own priorities within the given budget level. The scattering of political responsibility in times of budget cuts has been described earlier as an effect that occurs in the face of indecisiveness by principal (Posner & Blöndal 2012: 29). Using a more pragmatic approach, the central control philosophy relies upon simplicity and consistence nowadays using PI only when useful to support the internal dialogue and learning processes.

As showed earlier the budgetary process of the U.S. Forest Service seems reasonably well prepared for dealing with continuous uncertainty about authorized allocation. Nonetheless years of underfunding of output targets have had their impact on the organization. This became clear when the Forest Service profited from temporary extra funding from the federal recovery program (ARRA). Due to years of underfunding of actual targets, respondents indicated this extra money was spent at the forest level more or less in line with the squeaky wheel metaphor on things that required immediate attention resulting from long overdue maintenance.

The decreasing relevance of PI in budget preparation is not limited to outputs but also goes for policy outcomes. For example a Department of Transportation respondent noticed that DOT performance plans hardly received any attention from Congress due to the broken budget process. This way the budgeting process itself is undermining the performance budgeting process. With the analysis of Chapter 1 in mind this may be a less surprising observation as the national political arena is not the place to look for PB success in the first place.

Conclusion on impact crisis
Developing performance informed budget cuts as an alternative to across-the-board cuts and rationally re-prioritizing budgets in the face of budgetary insecurity were observed in the first two categories of responses. This supports the idea that PI will be used more intensively for budget reallocation as budgets becomes tighter and seems consistent with the observation over 30 years ago by Levine et al. In a study of how local governments manage fiscal stress they observed that the deeper a decline in revenue and the longer this lasts, the higher the likeliness that budget cuts will be targeted and programs will be terminated (Levine, Rubin & Wolohojian 1981: 194). On the other hand, the corroding of financial performance incentives and performance contracts that specify outputs, activities and costs was also observed in the case studies as well as declining interest in performance planning from political principals. These observations provide some evidence for the assumption that input steering takes over when a budget crisis hits.

In addition, the U.S. cases in particular illustrated how a PB system can function under pressure from a chaotic input oriented budgeting process recently seen in D.C. Although

these dynamics are observed in both countries, the U.S. sequestration seems to further deepen this problem of returning to input controls in the US. Installing mandatory cuts as a percentage of existing budget levels formalizes an input oriented approach towards budgeting that completely disregards performance or political priority setting altogether. In a budgeting climate like this, it is hard to claim that a PB approach would have superior results for an agency than the traditional approach to public sector budgeting as described by Wildavsky over 20 years ago (see Chapter 4.2). Moreover the apparently rational budgeting process based on realistic cost prices and workloads as developed by agencies throughout the years may have obscured a harsher reality of underfunded targets and more random allocation. This form of *squeaky wheel budgeting* represents a fundamental different approach to budgeting than realistic, transparent and predictable cost prices for public goods and services that are planned to achieve strategic policy objectives. It remains to be seen to what extent a well-developed PB system can be compatible with continuous reallocation process that characterizes a long period of fiscal consolidation.

Allen Schick once claimed that PB should not be viewed as the locomotive that drives government to change but rather as the caboose that confirms the transformations that have been made (Schick 2003: 102). The Great Recession raises the question how durable even these transformations are, potentially further diminishing the value of PB as a reform. Speculating somewhat further if, apart from favorable institutional agency features that this research focused on, an additional factor for PB to work consists of favorable fiscal conditions, the working of the PB recipe in terms of gains in efficiency and effectiveness, may be even more limited.

REFERENCES

Adler, Bill and Bill Jr. (1998). *The Reagan Wit – The Humor of the American President*, New York: William Morrow and Company, Inc.

Argyris, C (1994). *Good Communication That Blocks Learning.* Harvard Business Review July-August 1994.

Beckers, Victor and Ringeling, Arthur (2003). *Vragen over beleid, Perspectieven op waardering.* Utrecht: Lemma BV

Behn, Bob (2003). *Why Measure Performance? Different Purposes Require Different Measures.* Public Administration Review, Vol.63, Nr.5. pp 586-606

Bell S. (2002), *Institutionalism: Old and New.* In D. Woodward (Ed.) in *Government, Politics, Power and Policy in Australia* (7th ed.) Melbourne: Longman.

Bell S. (2011). *Do We Really Need a New 'Constructivist Institutionalism' to Explain Institutional Change?* British Journal of Political Science, Volume 41, No. 04 (October 2011), pp 883-906. DOI: http://dx.doi.org/10.1017/S0007123411000147

Blatter, J. and Haverland, M. (2014) *Designing Case Studies – Explanatory Approaches in Small-N Research*, Palgrave Macmillan

Bogt, Hendrik Jan Ter (1997). Neo-institutionele economie, management control *en verzelfstandiging van overheidsorganisaties - overwegingen voor verzelfstandiging en effecten op efficiëntie en financieel-economische sturing*. Capelle aan de IJssel: Labyrint Publication

Brady, Kevin (2008). *Forecasting and reporting performance - the search for the Holy Grail?* Policy Quarterly 4, 15-19

Brass, Clinton T. et al. (2006). *Congress and Program Evaluation*: *An Overview of Randomized Controlled Trials (RCTs) and Related Issues*. CSR Report for Congress.

Breul, Jonathan D. and Carl Moravitz (eds. 2007). *Integrating Performance and Budgets-The Budget Office of Tomorrow.* IBM Center for the Business of Government Book Series, Lanham, MD: Rowman & Littlefield

Broadnax, W.D., &. Conway, K.J. (2001). *The Social Security Administration and performance management*. In D.W. Forsythe (Ed.), Quicker, Better, Cheaper: Managing Performance in American Government (143-76). Albany: Rockefeller Institute Press

Chapman, Jeff and Duncan, Grant (2007). *Is there now a new New Zealand model?* Public Management Review, 9:1, p-1-25

Congiano, Marco, Teresa Curristine and Michel Lazare (eds. 2013), *Public Financial Management and its Emerging Architecture.* IMF Publication

Conlan, Timothy J., David R. Beam and Paul L. Posner (2002), *Pathways of Power The Dynamics of National Policymaking*. Paper prepared for the French Center on the United States.

Curristine, Teresa (2005), OECD Countries' Experience of Performance Budgeting and Management: Lessons Learned. In F. Stevens Redburn, Robert J. Shea and Terry F. Buss (Eds.), *Performance Management and Budgeting – How Governments can Learn from Experience,* (NAPA series on Transformational trends in Governance and Democracy), Armonk NY: M.E. Sharpe

Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). *Toward a stewardship theory of management*. Academy of Management Review, 22(1), 20-47

De Jong, M. , Van Beek, I. and Posthumus R. (2013), *Introducing accountable budgeting: lessons from a decade of performance-based budgeting in the Netherlands*, OECD Journal on Budgeting, Vol. 12/3

De Lancer Julnes, P. (2008). Performance Measurement beyond Instrumental Use. In Van Dooren W. & S. Van de Walle (Eds.), *Performance Information in the public sector: how it is used.* (pp. 58-71). Basingstoke: Palgrave McMillan.

De Vries, J. and Bestebreur, T. (2010). Budget reforms in the Netherlnds: Sadder but much wiser now. In Wanna, John, Jensen, Lotte and De Vries, Jouke (2010-eds). *The Reality of Budgetary Reform in OECD Nations – Trajectories and Consequences,* Cheltenham, UK and Northampton, MA USA: Edward Elgar, pp.221-39

Diamond, J. (2005). *Establishing a Performance Management Framework for Government*, IMF working paper.

Diamond, J (2011). *The Non-Technical Context of PFM Reform,* IMF working paper.

Dixit, A. (2002). *Incentives and Organizations in the Public Sector: An Interpretative Review*. The Journal of Human Resources, 37(4), 696–727.

Dull, Matthew (2009a), *Results-Model Reform Leadership: Questions of Credible Commitment*, Journal of Public Administration Research and Theory: J-PART 19(2): 255-84. Oxford University Press

Dull, Matthew (2009b), *De-PART: Perspectives on PART's Mixed Legacies,* Paper prepared for 31st annual APPAM Research Conference, November 5-7, 2009 – Washington D.C.

Egeberg, M. and Trondal, J. (2009), *Political Leadership and Bureaucratic Autonomy: Effects of Agencification*. Governance, 22: 673–688

Gilmour, John B. and Lewis, David E. (2006). *Does Performance Budgeting Work? An Examination of OMB's PART Scores.* Public Administration Review, Vol. 66, p.742-752

Gerrish, E. (2015), *The Impact of Performance Management on Performance in Public Organizations: A Systematic Review and Meta-Analysis,* Public Administration Review, Vol.75, Nr.5

Gore, Al (1995). Common Sense Government: Works Better & Costs Less. Third Report *of The National Performance Review*. Washington D.C.: U.S. Government Printing Office

Government Accountability Office (2004a). *Results Oriented Government: GPRA has Established a Solid Foundation for Achieving Greater Results.* GAO 04-38, a Report to Congressional Requesters

Government Accountability Office (2004b). *Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget*. GAO-04-174, a Report to Congressional Requesters

Government Accountability Office (2014). *Managing for Results – Agencies'Trends in the Use of Performance Information to Make Decisions.* GAO 14-747, a Report to Congressional Addressees

Frisco, Velda and Stalebrink, Odd J (2008). *Congressional Use of the Program Assessment Rating Tool.* Journal of Public Budgeting & Finance, Summer 2008- pp. 1–22

Hall, Peter A. and Taylor, Rosemary C.R. (1996). *Political Science and the Three New Institutionalisms*. MPIFG Discussion Paper 96/6, ISSN 0944-2073. Köln: Max Planck-Institut für Gesellschaftsforschung

Hammerschmid, Gerhard, Steven van de Walle and Vic Stimac (2013). *Internal and External Use of Performance Information in Public Organizations: Results from an International Survey*. Public Money and Management, July 2013

Hatry, Harry, Morley, Elaine, Rossman, Shelli, and Wholey, Joseph (2003). *How federal Programs Use Outcome Information: Opportunities for Federal Managers.* In Kamensky, John M. and Albert Morales (Eds.), *Managing for Results* Washington DC: IBM for the Business of Government

Hatry, Harry (2008). Results Based Budgeting. In Redburn, F. Stevens, Shea, Robert and Buss, Terry F. (2008 - eds). *Performance Management and Budgeting – How Governments can Learn from Experience,* (NAPA series on Transformational trends in Governance and Democracy), Armonk NY: M.E. Sharpe. pp. 293-318

Hay, C. (2006) Constructivist Institutionalism. In Rhodes, R. A. W., Binder, S. & Rockman, B. A. (Eds.) *The Oxford Handbook of Political Institutions* (Oxford: Oxford University Press) 56-74

Heckman, J., Heinrich, C., & Smith, J. (1997). *Assessing the performance of performance standards in public bureaucracies*. The American Economic Review, 87(2), 389-395.

Holmstrom, B., & Milgrom, P. (1991). *Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design.* Journal of Law, Economics, & Organization, 7, 24-52.

Immergut, Ellen M. (1998). *The Theoretical Core of the New Institutionalism*. Politics and Society, Vol. 26 – No. 1, p.4-35. Sage Publications Inc.

In 't Veld, R.J. (1998), The dynamics of policy instruments. In Peters, Guy B and Van Nispen, Frans K.M. (1998 -eds). *Public Policy Instruments – Evaluating the Tools of Public Administration.* Cheltenham, UK and Northampton, MA USA: Edward Elgar

IOFEZ, Interministerial consultations for financial and economic affairs (2004). *Policy budgets and policy accountability: evaluation – lessons from practice* (Self Evaluation of NL performance budgeting effort by Ministries)

Jansen, Paul & André de Waal (2004). *Prestatiemanagement van betrokkenheid in de publieke sector.* published in Gedrag & organisatie, Vol 17 – No 6, december 2004

Jenkins, Simon (2006), *Thatcher and Sons – a Revolution in Three Acts,* London: Allen Lane

Joyce, Philip G. (2003). *Linking Performance and Budgeting: Opportunities in the Federal Budget Process.* In Breul, Johnathan D. and Carl Moravitz (Eds.) Integrating Perfromance and Budgets, Washington DC: IBM for the Business of Government

Kaufman, Herbert (1960). *The Forest ranger – a study in administrative behavior.* Republished by Resources for the Future, Washington DC in 2006

Kersbergen Kees Van, Waarden, Frans Van (2004). *'Governance' as a bridge between Disciplines: Cross-disciplinary inspiration regarding shifts in governance and problems of governability, accountability and legitimacy*. European Journal of Political Research No. 43, p.143-171. Oxford UK: Blackwell Publishing Ltd.

Kim, John M. (2010). Korea's four main budgetary reforms: catching up with a big bang. In Wanna, John, Jensen, Lotte and De Vries, Jouke (2010-eds). *The Reality of Budgetary Reform in OECD Nations – Trajectories and Consequences,* Cheltenham, UK and Northampton, MA USA: Edward Elgar, pp.221-39

Kraan, D.J. (2011), "Building on Basics – OECD Value for Money Study, Final Report: Sections from Chapter 1 (Introduction on new trends in public administration)", hand-out at the 7th annual meeting on performance and results, OECD, Paris, 9-10 November, www.oecd.org/gov/budgeting/49042446.pdf

Kroll, Alexander (2013). *The Other Type of Performance Information: Nonroutine Feedback, Its Relevance and Use,* Public Administration Review Volume 73, Issue 2, pages 265–276, March/April 2013

Kroll, Alexander (2015). *Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research*, Public Performance and Management Review Vol. 38 - No. 3 pp. 459-486.

Laegreid, P., Roness, P. G., & Rubecksen, K. (2008). Performance Information and performance steering; integrated system or loose coupling. In Van Dooren W. & S. Van de Walle (Eds.), *Performance Information in the public sector: how it is used.* (pp. 125-140). Basingstoke: Palgrave McMillan.

Levine, C.H, I.S. Rubin & G.G. Wolohojian (1981), *The politics of retrenchment: how local governments manage fiscal stress*, Sage, Beverly Hills

Lindblom, Charles (1959) *The Science of Muddling Through*, Public Administration Review 19: p. 79-88

Linsenmann, Ingo, Meyer, Christoph O. and Wessels, Wolfgang (eds. - 2007). *Economic Government of the EU - A Balance Sheet of New Modes of Policy Coordination.* P algrave Macmillan

Mahler, Julianne (1997) *Influences of Organizational Culture on Learning in Public Agencies,* Journal of Public Administration Research and Theory: J-PART 7(4): 519-540. Oxford University Press

Marcel, Mario (2013). *Budgeting for Fiscal Space and Government Performance beyond the Great Recession.* Paper presented at the fifth annual meeting of OECD parliamentary budget officials and independent fiscal institutions in Ottawa, Canada, Feb 21-22, 2013

March, James G. and Olsen, Johan P. (1984). New Institutionalism: *Organizational Factors of Political Life*. The American Political Science Review, Vol. 78 – No.3, p.734-749

March, James G. and Olsen, Johan P. (1998). *The Institutional Dynamics of International Political Orders*, International Organization, Vol. 52 – No.4, p.943-969. Cambridge University Press, UK

Melkers, J. and Willoughby, K. (2004): S*taying the Course: The Use of Performance Measurement in State Governments*. In Breul, Johnathan D. and Carl Moravitz (Eds.) Integrating Perfromance and Budgets*,* Washington DC: IBM for the Business of Government

Mills, Russel W. 2013, *Incident Reporting Systems – Lessons from the Federal Aviaition Administration's Air Traffic Control Organization*. Publication of IBM Center for the Business of Government, Improving performance series.

Minister van Financiën (2011), *De baten en lasten van het baten-lastendienstmodel. Evaluatie Regeling baten-lastendiensten 2007.* Tweede Kamer, 2010-2011, 28737 nr. 21.

Moynihan, Donald P. (2005). *Goal-Based Learning and the Future of Performance Management.* Public Administration Review Volume 65, No 2

Moynihan, Donald P. (2008). *The Dynamics of Performance Management – Constructing Information and Reform*. Washington D.C.: Georgetown University Press

Moynihan, Donald P., and Daniel P. Hawes (2012). *Responsiveness to reform values: The influence of the environment on performance information use.* Public Administration Review 72.s1 (2012): S95-S105.

Moynihan, Donald P., and Noel Landuyt. 2009. '*How Do Public Organizations Learn? Bridging Cultural and Structural Perspectives'*. Public Administration Review 69 (6): 1097–1105.

Moynihan, Donald P. and Pandey, Sanjay K. (2010). *The Big Question for Performance Management: Why Do Managers Use Performance Information*?, Journal of Public Administration and Theory Advance Access, Oxford University Press.

Moynihan, Donald P. and Lavertu Stéphane (2012). *Does Involvement in Performance management Routines Encourage Performance Information Use? Evaluating GPRA and PART.* Public Administration Review 74(4): 592-602 ASPA

Moynihan, Donald P. and Alexander Kroll, (2015) *Performance Management Routines That Work? An Early Assessment of the GPRA Modernization Act*. Public Administration Review (2015) forthcoming

Niskanen, W.A. (1971), *Bureaucracy and Representative Government*, Chicago: Aldine Atherton Inc.

Niskanen, W.A. (1973), *Bureaucracy: Servant or Master? - Lessons from America*, Londen: The Institute of Economic Affairs

OECD (2007), *Performance Budgeting in OECD Countries*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264034051-en

OECD (2014), *Note on the Evolving Role of Performance and Results.* Handout for session during 35th OECD Senior Budget Officials Meeting, Berlin Germany, June 12-12 2014

OECD (2015), *Building on Basics, Value for Money in Government,* OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264235052-en

Olson, Mancur (1965). *The Logic of Collective Action – Public Goods and the Theory of Groups*. Cambridge MS: Harvard University Press

Olson, Mancur (1973). Evaluating performance in the Public Sector. In Moss, Milton (Ed.) *The Measurement of Economics and Social Performance.* NBER publication pp.355-409

Ostrom, Elinor (1986). *An Agenda for the Study of Institutions*. Public Choice Nr. 48, p.3-25. Dordrecht: Martinus Nijhoff Publishers.

Osborne, David and Plastrik, Peter (1997). *Banishing Bureaucracy: The Five Strategies for Reinventing Government*, Reading MS: Addison Wesley

Pan, Wen (2014). *Crises and Opportunities – Strengthening European Union Economic Governance after the 2008 Financial Crisis,* Ridderprint BV, The Netherlands

Partnership for public service / IBM Center the Business of Government 2011. *From Data to Decisions – the Power of Analytics*

Perry, James L. & Annie Hondeghem (2008). *Building Theory and Empirical Evidence About Public Service Motivation*. International Public Management Journal 11(1): 3-12

Pierson Paul (2000). *Increasing Returns, Path Dependence, and the Study of Politics.* The American Political Science Review, Volume 94, Issue 2, p.251-67

Pollitt, Christopher, Colin Talbot, Janice Caulfield and Amanda Smullen (2004). *Agencies – How governments do Things Through Semi-Autonomous Organizations.* Basingstoke (UK) and New York: Palgrave Macmillan.

Pollitt, Christopher (2006), *Performance Management in Practice: A Comparative Study of Executive Agencies,* Journal of Public Administration and Research and Theory Vol. 16 – Issue 1, pp 25–44

Posner, Paul L. (2004). *Another New Politics of Budgeting: The Pathways Framework*. Annual Meeting of Association for Budgeting and Financial Management, Chicago, IL, October.

Posner, Paul L. and Chung-Keun Park (2008). *Role of the Legislature in the Budget Process: Recent Trends and Innovations*, OECD Journal on Budgeting Vol. 7 – No.3

Posner, Paul L. and Fantone, Denise M. (2008). Performance Budgeting – Prospects for Sustainability. In Redburn, F. Stevens, Shea, Robert and Buss, Terry F. (2008 - eds). *Performance Management and Budgeting – How Governments can Learn from Experience,* (NAPA series on Transformational trends in Governance and Democracy), Armonk NY: M.E. Sharpe. pp. 92-13

Posner, Paul L. (2009). *Performance Budgeting: Informing Hard Choices Facing the Nation.* Statement Prepared for Hearing before the Senate Budget Committee October 29, 2009

Posner, Paul L. and Jón Blöndal (2012), *Democracies and Deficits: Prospects for Fiscal Responsibility in Democratic Nations*. Governance, Volume 25, Issue 1, p. 11–34, January 2012

Posner, Paul L. and Mahler, J. (2012). *Performance Movement at a Crossroads: Information, Accountability and Learning,* paper presented at 8th Transatlantic Dialogue at University in Nijmegen, June 7th 2012

Radin, Beryl A. (2006). *Challenging the Performance Movement – Accountability, Complexity And Democratic Values.* Washington D.C.: Georgetown University Press

Redburn, F. Stevens, Shea, Robert and Buss, Terry F. (2008 - eds). *Performance Management and Budgeting – How Governments can Learn from Experience,* (NAPA series on Transformational trends in Governance and Democracy), Armonk NY: M.E. Sharpe

Rhodes, R.A.W. (1997). *Understanding Governance – Policy Networks, Governance, Reflexivity and Accountability*. Buckingham, PA: Open University Press

Robson C. 1993. *Real World Research: A Resource for Social Scientists and Practitioners-Researchers*. Oxford: Blackwell

Robinson, Marc & Brumby, Jim (2005). *Does Performance Budgeting Work? An Analytical Review of the Empirical Literature.* IMF Working Paper WP/05/210

Robinson, Marc (2014). *Connecting Evaluation and Budgeting,* ECD Working Paper Series No. 30. published by the Independent Evaluation Group, The World Bank Group, Washington D.C.

Schick, A (1993) *Governments versus Budget Deficits.* In Weaver, R. Kent & Rockman, Bert A. (1993- eds), *Do Institutions Matter? – Government Capabilities in the United States and Abroad*. Washington D.C.: The Brookings Institution

Schick, A. (2003) *The Performing State: reflection of an idea whose time has come but whose implementation has not.* OECD Journal on Budgeting, Volume 3 - No.2.

Schick, A. (2014) *The metamorphoses of performance budgeting.* OECD Journal on Budgeting, Volume 2013 – No.2

Schmidt, Vivien A. (2008), *Discursive Institutionalism: The Explanatory Power of Ideas and Discourse,* Annual Review of Political Science Vol. 11: 303-326 (June 2008). First published online as a Review in Advance on January 22, 2008 DOI:10.1146/annurev.polisci.11.060606.135342

Schoch M. & C. den Broeder (2013), *Linking Information on Policy Effectiveness to Budget Decisions in the Netherlands,* OECD Journal on Budgeting, Vol. 12/3

Scott, G. (2003). *The Learning Government*. OECD Journal on budgeting, Volume 3 – No.2.

Senge, P.M. (1990). *The Fifth Discipline- The Art and Practice of the Learning Organization.* New York: 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

Ten Heuvelhof, E.F. (1999) Van Sturen naar kansen benutten. In In 't Veld, R.J. (1999-red.) *Sturingswaan & ontnuchtering – het karakter van overheidssturing verandert. Het is zelfs de vraag in hoeverre de term sturing nog op zijn plaats is*. Utrecht: Lemma BV, pp. 70-85

Tipple, Terence J. and Wellman, Douglas J. (1991) *Herbert Kaufman's Forest Ranger Thirty Years Later: From Simplicity and Homogeneity to Complexity and Diversity.* Public Administration Review 51(5):. 421-428

True, James L., Jones, Bryan D. and Baumgartner, Frank R. (1999). *Punctuated Equilibrium Theory – Explaining Stability and Change in Public Policymaking*

Van Dooren, Wouter and van de Walle, Steven (2008 - eds). *Performance Information in the Public Sector – How it is used.* Basingstoke: Palgrave MacMillan

Van Dooren, W. (2011). *Better Performance Management – Single- and Double Loop Strategies.* Public Performance & Management Review, Vol.34, No. 3, p.421-434. March 2011.

Kersbergen Kees Van, Waarden, Frans Van (2004). *'Governance' as a bridge between Disciplines: Cross-disciplinary inspiration regarding shifts in governance and problems of governability, accountability and legitimacy*. European Journal of Political Research No. 43, pp.143-171. Oxford UK: Blackwell Publishing Ltd.

Van der Knaap, Peter (1997). *Lerende overheid, intelligent beleid – de lessen van beleidsevaluatie en beleidsadvisering voor de structuurfondsen van de Europese Unie.* Den Haag: Phaedrus.

Van Dooren, W (2011), *Better Performance Management – Some Single and Double Loop Strategies*, Public Performance and Management Review, Vol. 34 – No 3, pp. 421-434

Van Nispen, Frans K.M. (1993). *Het dossier heroverweging,* Delft: Eburon

Van Nispen, Frans K.M. & Johan J.A. Posseth (2006), *Performance Budgeting in the Netherlands: Beyond Arithmetic.* OECD Journal on Budgeting, Vol. 6, No. 4, pp. 37-62.

Van Nispen, Frans K.M. (2015), *Policy Analysis in Times of Austerity: A Cross-National Comparison of Spending Reviews* (forthcoming)

Van Thiel, S. (2008). *The empty nest syndrome: Dutch ministries after the separation of policy and administration*. Paper presented at IRSPM 12, 26-28 March 2008, Brisbane Australia.

Verhoest, Koen, Van Thiel, Sandra, Bouckaert, Geert and Laegreid, Per (2011). *Governing Public Agencies in the 21st Century – International lessons and policy recommendations by the COST Action IS0601 on Comparative Research into Current Trends in Public Sector Organizations –CRIPO/COBRA.*

Verhoest, Koen, Van Thiel, Sandra, Bouckaert, Geert and Laegreid, Per (2012). *Government Agencies – Practices and Lessons from 30 Countries.* Basingstoke (UK): Palgrave Macmillan.

Von Meyenfeldt, Lone, Schrijvershof, Carlien and Wilms, Peter (2008). *Tussenevaluatie Beleidsdoorlichting.* Den Haag. Evaluation study conducted by Aarts De Jong Wilms Goudriaan Public Economics bv (APE) for the NL Ministry of Finance

Walker, Richard M., Chan Su Jung and Boyne, George A. (2013). *Marching to Different Drummers? The Performance Effects of Alignment between Political and Managerial Perceptions of Performance.* ManagementPublic Administration Review, Vol. 73, Iss. 6, pp. 833–844.

Weaver, R. Kent & Rockman, Bert A. (1993- eds), *Do Institutions Matter? – Government Capabilities in the United States and Abroad*. Washington D.C.: The Brookings Institution

Wildavsky, Aaron (1961). *Political Implications of Budgetary Reform*, Public Administration Review No. 21

Wildavsky Aaron (1969), *Rescuing Policy Analysis from PPBS*, Public Administration Review, Vol. 29, No. 2 (Mar. - Apr., 1969), pp. 189-202, Blackwell Publishing on behalf of the American Society for Public Administration

Wildavsky, Aaron (1992). *The New Politics of the Budgetary Process*, New York: HarperCollins Publishers Inc.

Wilson, James Q. 1989. *Bureaucracy. What Government Agencies Do and Why They Do It*. New York: Basic Books

Yin, Robert K. 2003. *Case Study Research: Design and Methods,* Thousand Oaks: Sage

APPENDIX I - BUDGET PROGRAMS NETHERLANDS MINISTRY OF AGRICULTURE FOR FUNDING SBB

Article 23 Nature

Central Government seeks to ensure the biodiversity and the increasing quality of our living environment. The decline in the wealth of plants, animals and the ecosystems has to be stopped, as these functions would otherwise be lost. Biodiversity has many functions, including meeting recreational needs and providing an attractive living environment and climate for establishing businesses. Conservation of biodiversity forms part of the international agreements as laid down in the Biodiversity Convention 1992. Biodiversity conservation therefore has a central place in the nature policies focused on the 2010 objective, stopping the further decline of all kinds of wealth.

23.12 Acquisition Ecological Basic Structure

Background Regarding the realization of the EHS, the Ministry of Agriculture, Nature and Food Quality makes available financial resources to purchase land on behalf of the area management organizations.

23.13 Management Ecological Basic Structure

Background Management within a related network of existing nature, the allocated Nature Conservation Areas and 18 National Parks. An important ancillary management objective is to comply with recreational demands and opening up nature managed by private individuals and private nature conservation organizations, unless this entails risks to special species or ecosystems. The management provides adequate opportunities and guarantees for the protection of endangered species that are dependent on the managed areas (in particular the species listed in the Multi-year Implementation Species Policy Program).

23.14 Management outside the EHS and protection of international biodiversity

National

- Nature management, agricultural nature management and landscape management.
- Performance of the international obligation to designate one or more nationally or internationally important ecosystems as a national park.
- Providing sustainable protection to all target species of the species policy.
- Subsidizing the LNV share in the costs of acquiring existing and new nature (the subsidy scheme for private area management nature conservation organizations.

International

- Performance of the international obligations intended to stop the decline of biodiversity across the world.

254

- Contributing to worldwide action for sustainable development and conservation of biodiversity.
- Identifying the PAN-European-Ecological Network.

Article nr. 24 Landscape and Recreation

Conservation and development of landscape and making the Netherlands attractive for recreation.

24.11 National Landscapes

Motives Conservation, management and consolidation of the unique landscape, cultural-historical and natural qualities of 20 National Landscapes.

24.12 Landscape – general

Background The Central Government has handed over to the provinces responsibility for the basic quality of the landscape. This is defined as the conservation and consolidation of the natural, cultural, user and perceived quality of the landscape. In this connection the Central Government has a facilitating and stimulating role. An example of this is the part played by the Central Government landscape advisor. In order to provide further support to provinces and local authorities the Central Government drew up the Landscape Quality Agenda Guide

24.14 Recreation – general

Background The Central Government wants a high-quality level of facilities in recreation areas and, to this end, is financing the recreational management tasks of the National Forest Service and the Middle Delfland recreational amenities board. The Central Government is also creating possibilities for entrepreneurship in the field of recreation. It also wants to link 492 hectares of regional greenery to major residential building locations in 2010. Important here is that these areas are fully open to the public and that entry is free of charge.

FORMAT USED FOR SEMI-STRUCTURED INTERVIEWS AND INTERVIEWS SCHEDULE

Variable	Indicator	Result			Quotes
		Absent	*Present to some extent*	*Clearly present*	
PB 1.1 - PB is used by the AGENCY in addition to traditional budgeting					
PB 1.2 - PB is used by the AGENCY's principal to control the AGENCY	PB is expected to be a major influence in the relation of the AGENCY with its principal if (on a mounting scale): • There is clear alignment of the information monitored and the principal's formal goals (1 pt) • Monitoring data are used frequently in a dialogue between AGENCY and principal (2 pt). • This dialogue results in concrete consequences and actions (3 pt) • There are provisions for financial incentives related to performance (4pt) • These provisions are actually used when	0, 1 pt	2,3 pt	4,5 pt	

256

	targets aren't met (5 pt)			
PB 2.1 - A high degree of de jure PB implementation	_Direct or formula performance budgeting._ (1 pt) In some cases a direct link between performance results and resource allocation and accountability is in place. _Performance informed budgeting_ (2 pt). More often the link is indirect and planned performance targets and results are used for planning and accountability purposes only. _Presentational performance budgeting_ (1 pt). PB systems that have no link between performance information and funding and performance information is used for accountability only. What is _the frequency of performance measurement and reporting?_ This may be done only when the yearly budget or report comes around (0 pt) or as part of a monitoring cycle with a higher frequency (1 pt).	0, 1 pt	2,3 pt	4,5 ,pt	

		0, 1 pt	2,3 pt	4,5 pt	
	Are there *measures taken to ensure the quality and impartiality* of the reported performance information? Examples are organizational checks and balances or incidental or systematic auditing. If so 1 pt, if not 0 pt				
PB 2.2 - A high degree of de facto PB implementation	On a mounting scale de facto PB implementation can be characterized by: ✓ Reporting of the required performance information in a complete and timely manner (1pt) ✓ Implicit reference to performance information by adopting performance lingo in communication (2 pt). ✓ Explicit reference to performance information in communication and decision making (3 pt) ✓ Decision-making matches formal conclusions and recommendations	0, 1 pt	2,3 pt	4,5 pt	

	based on expected or demonstrated performance (4 pt) ✓ Budgetary consequences are linked to performance based decisions (5 pt) >= 3 pt only If substantiated by example				
H 1.1 - Has there been a critical juncture in the accountability chain or policy field?					
H 1.2 - If so, has this critical juncture lead to a broadly felt problem to which PB was seen as a solution.					
H 2.1 - An internal advocate or external champion of PB in a powerful position.				Relatively Powerful position FEZ Or other PB Champion	
H 2.2 - A policy field in which specialists are dominant, relative to the other political actors.				Policy field with relatively dominant expert pathway	
S 1.1 –<u>Senge's learning</u>					

disabilities lacking				
S 1.2 - Participative openness: Important issues are being discussed openly and fairly				
S 1.3 - Reflective openness:The ability to continually challenge one's own thinking				
S 2.1 - Shared view of the meaning of measured performance data expressed by 100% fit between the formal performance measures and the PB system				
CX 1.1 - Macro economic factors				
CX 2.1 - Shifting political preferences				
CX 3.1 - Others				

SBB (National Forest Service of the Netherlands): Interviews Dec 2011 – Feb 2012
Ministry of Economic Affairs, Sr Advisor Forestry, Nature Department
Ministry of Economic Affairs, Advisor Forestry, Financial and Economic Affairs Dept.
SBB, Deputy Director (HQ)
SBB, Controller (HQ)
SBB, Specialist – Fauna - ecology (HQ)
SBB, Sr. ecologist, inventarization and data management specialist (HQ)
SBB, Director Services Unit
SBB, Director Eastern Region
SBB, Controller Eastern Region
SBB, Chief Forester within Eastern Region
LVNL (Air Traffic Control Netherlands): Interviews March – November 2012
Ministry of Infrastructure, DGB, Sr. Advisor Aviation
Ministry of Infrastructure, DGB, Director Aviation
LVNL, Director Corporate Development
LVNL, Manager Strategy
LVNL, Manager Performance
LVNL, General Manager Regional Unit
LVNL, General Manager Strategy and Performance
LVNL, General Manager Operations
LVNL, Chief Financial Officer
LVNL, Manager Quality Assurance
U.S. Forest Service: Interviews Sept-Oct 2012 & June-July 2013
Branch Chief, Performance Management (national level)
Regional Strategic Budget Coordinator Region 6 (R6)
Analyst Economist R6
Budget Analyst R6
Deputy Forester of NF with R6
Budget specialist NF within R6
Silviculturist responsible for timber program NF within R6
Ecologist responsible for biodiversity programs   NF within R6
Federal Aviation Administration: Interviews Oct 2012 & July 2013
Strategic Planning Coordinator, Dept. of Transportation
Manager Strategic Planning, Office of Aviation Policy and Plans (OAP&P)
Special Advisor for Planning, OAP&P
Program Analyst, OAP&P
Program Analyst, OAP&P
Management and Program Analyst, Office of Budget
Executive Director Air Traffic Organization (ATO)
Acting Director Strategic Enterprise Directorate ATO
Other interviews on general issues US cases (not using semi structured interview format)
Shelley Metzenbaum, *Associate Director for Performance and Personnel Management (OMB)*
Mark Bussow, *Performance Manager at Office of Management and Budget*
Academics consulted on this study and preliminary results (list is not conclusive)
Bob Behn, *Sr. Lecturer in Public Policy, Harvard Kennedy School*

Russ Mills, *Asst. Professor Pol. Science, Bowling Green State University Ohio (and formerly at FAA)*
Aimee Franklin, *Professor in Political Science Department, University of Oklahoma*
Matt Dull, *Associate Professor Virginia Tech University*
Wouter van Dooren, *Associate Professor Public Adm. & Management, Antwerpen University*

In November/December 2007 I conducted an independent study at George Mason University of the White House's Program Assessment Rating Tool (PART). This study, which was supported by the Netherlands Ministry of Finance and generously hosted by GMU's Paul Posner, proved to be pivotal to this study in several respects. It helped me gain familiarity with the particular characteristics of the U.S. Federal Government system compared to the Netherlands as well as with U.S. federal performance management and performance budgeting initiatives. This knowledge proved to be pivotal for the problem analysis leading to this study as well as for assessing the U.S. case studies. In 2007 the following people were interviewed (their functions obviously reflect those at the time)
Paul L. Posner, *Professor, Director of MPA Program, Public and International Affairs (Until 2005 Professor Posner was Managing Director. Federal Budget Analysis and Intergovernmental Relations at GAO)*
Jonathan D. Breul*, Executive Director IBM Center for the Business of Government. (Mr. Breul served earlier as a sr. advisor at OMB)*
Philip G. Joyce, *Professor of Public Policy and Public Administration at School of Public Policy and Public Administration of George Washington University*
Robert J. Shea, *Associate director of administration and government performance at the President's Office of Management and Budget. (Mr. Shea headed the president's budget and performance integration initiative and oversaw PART implementation and use by OMB)*
Maurice P. McTigue, *Director Government Accountability Project and Distinguished Visiting Fellow at Mercatus Center, George Mason University (Hon. McTigue served as a member of parliament and a cabinet minister in New Zealand in the 1980s and 1990s)*
Eileen Norcross, *Senior Research Fellow, Government Accountability Project at Mercatus Center, George Mason University*
Kyle M. McDonald, *Program Associate, Government Accountability Project at Mercatus Center, George Mason University*
Denise M. Fantone, *Acting Director Health Care at Government Accountability Office (GAO)*
Jacqueline M. Nowicki, *Assisting Director Strategic Studies at GAO*
Matthew Dull, *Assistant Professor, Virginia Tech's Center for Public Administration and Policy*
Clinton T. Brass, *Analyst in American National Government, Government and Finance Division, Congressional Research Service*
Jeffrey Newsome, *Director of Program Performance at the Budget Office of the US Department of Transportation*
Roald van der Linde, *Counselor Finance Department NL Embassy in Washington D.C.*
Dale Morris, *Sr. Advisor Economic Affairs NL Embassy in Washington D.C. (Mr. Morris is a former budget aide to a member of Congress)*

262

APPENDIX III   QUESTIONNAIRE (FAA/ATO VERSION)

Thank you for participating in this survey! Before starting the 17 questions, a few announcements on the survey itself:

PURPOSE AND CONFIDENTIALITY
- This questionnaire will be used for a scientific study that is conducted fully independent from any executive office or interest group.
- Results from this questionnaire will be used for academic purposes only.
- The answers will be treated with strict confidentiality to make sure that publicized data cannot be traced to an individual respondent at any time.

EXPLANATORY NOTES
- Unless stated otherwise, only one answer should be chosen for each question.
- The term 'performance measurement' as used in this survey, refers to measurement of results in the fields of work of FAA/ATO like for instance air traffic safety, delays, airport capacity, noise levels.
- The term 'performance information' as used in this survey, refers to the results of these measurements and include both performance indicators as quantitative analysis and evaluation.

GENERAL

1. What is your age?
a) <25
b) 26-40
c) 41-55
d) >55

2. What is your sex?
a) M
b) F

3. In which ATO functional area are you employed?
a) Central Management
b) Program Management Operations
c) Safety and Technical training
d) Operations

4. How many years have you been working for the FAA?
a) <2
b) 2-10
c) 10-25
d) >25

5. **What is your line of work?**
a) Management
b) Staff/Technical Support
c) Finance
d) Air traffic control operations
e) Other:………..

POLICY AND PERFORMANCE MEASUREMENT

6. **Which of these factors plays the largest role in helping shape FAA/ATO's policies (please prioritize these answers)**
a) Political ideals
b) Policy priorities from the White House and/or the DOT
c) Interests of large stakeholders in business and government
d) Interests as expressed by (groups of) citizens

7. **Which of these factors plays the largest role in helping shape FAA/ATO's policies (please prioritize these answers)**
a) Political debate
b) Media (TV, papers)
c) International professional standards
d) Technical developments

8. **Do you believe that using performance measurement has a positive influence on your work effectiveness?**
a) Yes
b) I am not sure
c) No

9. **Do you believe that using performance measurement has a positive influence on the efficiency of FAA/ATO as an organization?**
a) Yes
b) I am not sure
c) No

10. **How frequently do you use performance information in your work?**
a) Regularly
b) Occasionally
c) Hardly ever
d) Never

11. **When performance information is used, for what purpose is it used? (multiple answers may be chosen)**
a) Performance reporting for external accountability
b) Setting program priorities
c) Strategically reallocate internal resources

264

d) Understand the impact of external events on performance goals
e) Deciding on outsourcing decisions
f) Developing and managing contracts
g) Monitor cost and performance and contract management
h) Allocate funds to third parties
i) Coordinating program efforts with other internal or external organizations
j) Analyzing productivity and funding levels
k) Allocating internal funds
l) Identifying service problems and changing work processes
m) Adopting new program approaches following evaluation
n) Motivate staff to act consistent with organizational goals
o) Setting individual job expectations for staff
p) Rewarding staff

12. To what extent are you familiar with the performance indicators used for your work?
a) I know the performance indicators as well as the target values used
b) I know the performance indicators but no target values
c) I've heard something about them but not a lot
d) I am not familiar with them

13. Who primarily sets performance indicators within FAA/ATO?
a) Budget and Financial staff
b) Technical specialists from staff and operations
c) Answers a and b jointly
d) External parties outside FAA/ATO
e) Other:

CORPORATE CULTURE
Please react to the following statements by selecting the answer that you feel is most appropriate

14. Within FAA/ATO important issues are being discussed openly and fairly

I totally disagree	I disagree somewhat	I disagree nor agree	I agree somewhat	I totally agree

15. FAA/ATO is an organization that is able to continually question and challenge its own way of thinking

I totally disagree	I disagree somewhat	I disagree nor agree	I agree somewhat	I totally agree

16. If FAA/ATO is faced with problems, a good problem analysis is usually made before taking action.

I totally disagree	I disagree somewhat	I disagree nor agree	I agree somewhat	I totally agree

17. If things go wrong in FAA/ATO's field of work, lessons are usually learned as a result.

I totally disagree	I disagree somewhat	I disagree nor agree	I agree somewhat	I totally agree

APPENDIX IV   CASE SELECTION

As mentioned in CH 2.4 and 4.2 a suitable case for testing the model has to meet a number of criteria. These are summarized into 5 conditions that apply to potential case material:

1. **It involves a national policy goal and performance targets tied to the national budget**
2. **Policy execution takes place in principal-agent setting (preferably a National HQ and regional agents) with the (plausible) presence of the associated problems of:**
   - conflicting goals and
   - superior access to information on the part of the agent
3. **The program and its execution are regarded a PB success[27]**
4. **There is de facto performance steering on the part of the principal (as defined in CH.2):**
   - PB is used by the agency's principal to control the agency
5. **There is de facto performance management on the part of the agent (as defined in CH. 4)**
   - PB is used in addition to traditional budget preparation

In order to limit the number of independent variables, related programs in terms of policy fields were sought. For the Dutch case selection a shortlist was made way with the help of expert opinions. Because of easy access to a PB database on U.S. national programs, potential NL cases could be systematically matched with U.S. programs concerning similar policy fields. This resulted in a potential US and a potential NL case that cover a principal-agent relationship in a single policy field. Additional factors determining the final case selection are obviously the access to reliable information and the ability to establish contacts with relevant professionals using the professional network of the committee and the author.

Netherlands (NL) shortlist
A shortlist of NL cases was constructed based on a variety of sources. Several studies have been conducted on the steering relationship between the NL central government and its agencies. Some of these covered the entire NL central government while others focused on a selection of cases. The main sources used are:

- De Kruijf, Johan A.M., 2011. *Autonomy or compliance in public service provision – Legal and economic autonomy and democratic control on service provision by ZBO's*. Dissertation University Twente
- Evaluatie Agentschapsmodel, Ministerie van Financien 2011 (to be published)

---

[27] This condition is met if performance management and output funding occur to a relatively large extent according to the Ministry of Finance (or equivalent body) and/or independent experts.

- Van Meerkerk, Ingmar & Van Thiel, Sandra 2011. *De verborgen eigenaar: De financieel-economische sturingsrelatie tussen ministeries en zelfstandige bestuursorganen.* Published in Bestuurswetenschappen 2011-nr. 1
- Algemene Rekenkamer 2005, Interne notitie over prestatiesturing. (Author: Rudi Turksema et al.)

In addition to this literature study, the suitability for agencies in the Netherlands was discussed with Dr. Sandra van Thiel, Dr Rudi Turksema and the staff of the Minstry of Finance tasked with the evaluation of the NL agency steering model. This resulted in a shortlist of principal-agent relationships in which output bases funding and de facto performance management are claimed to have been dominant in recent years.

Table 1

NL Shortlist		
Principal	**Agent**	**Policy area/program**
Min van Justitie / V&J	Dienst Justitiële Inrichtingen (DJI)	Justice- detention
Min van Justitie / V&J	Integratie & Naturalisatiedienst (IND)	Immigration
Min van Justitie / V&J	Raden voor de Rechtsbijstand (RVR)	Justice
Min. van OCW	Vervangingsfonds/ Participatiefonds	Education/ social security
Min. van Financiën	Belastingdienst	Tax handling
Min. van VROM / BZK	Centraal Fonds Volkshuisvesting (CVF)	Housing
Min van VenW I&M	Rijksdienst voor het Wegverkeer (RDW)	Road transportation (car registration)
Min van VenW / I&M	Luchtverkeersleiding Nederland	Aviation (traffic control)
Min. van EZ/ELenI	Nederlandse Mededingings Autoriteit (NMA)	Oversight Economy
Min van EZ / ELenI	Senter Novem / Agentschap NL	Government Shared Service / handling of subsidies/ grants
Min van LNV / EL&I	Plantenkundige Denst (PD)	Plant disease control
Min van LNV / EL&I	Staatsbosbeheer	Forestry service
Min van LNV	Nederlandse Algemene Keuringsdienst (NAK)	Food safety

The highlighted relationships represents a subset of policy areas where public services are believed to be the least likely affected by differences in national context. For this reason these associated five policy areas were analyzed for matching U.S. cases.

Matching U.S. Programs

The level of documentation of PB aspects of U.S. federal programs allows for a systematic analysis of potential case-candidates. This goes in particular for conditions 3-5. Conditions 1 and 2 can be addressed reasonably adequately by examining the organization of the program on the agency's website. Potential cases do however need additional in depth analysis once selected. The questions A-G in Table 2 are to be answered for this further analysis. For now the presence of these conditions is assumed in all cases.

With regard to conditions 3-5, a major source for analyzing U.S. federal programs, are the program assessments made by the White House's Office of Management and Budget for the FY 2002-2008 budget proposals. Known as the Program Assesment Rating Tool (PART), OMB used a standardized set of (in most cases) 25 questions to measure program effectiveness of over 1000 federal programs. The answers to these questions determined if a program was rated as either *effective*, *moderately effective*, *adequate*, *inadequate* or *results not demonstrated.*

As a Proxy for conditions 3-5, eight PART question bear particular relevance. If these were answered with a YES by the OMB examiner, this indicates that the conditions sought are met for that particular program. Table 2 shows the PART questions and the conditions to which they refer. Two additional restrictions have been applied:

- A program that is overall rated as *ineffective* or *results not demonstrated* by OMB is not viewed as an appropriate candidate for representing a good practice in terms of PB.
- Financially less significant programs (< 25 million US$) have been excluded from the results

Table 2

A) Have policy objectives been communicated in the national budget/performance plan? **B)** Have these objectives been translated into program goals and performance indicators?	**Condition 1** National policy goal and performance targets tied to the national budget
C) Is execution of tasks performed by a decentralized government organization? **D)** Does a national ministry and/or agency headquarter act as a principal to the executive agencies?	**Condition 2** Principal-Agent setting with conflicting goals and superior access to information by agent

E) Are different offices (e.g. regional) tasked with identical or highly comparable duties? F) Can conflicting goals be identified and made plausible? G) Can information asymmetry be identified and made plausible?	
Nr. PART question	**Relevant for condition nr.**
2.4 Does the program have baselines and ambitious targets and timeframes for its annual measures?	
2.6 Are independent and quality evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?	**Condition 3** The program and its execution are regarded a PB success
2.8 Has the program taken meaningful steps to correct its strategic planning deficiencies?	
3.4 Does the program have procedures (e.g., competitive sourcing/cost comparisons, IT improvements, appropriate incentives) to measure and achieve efficiencies and cost effectiveness in program execution?	
2.5 Do all partners (including grantees, sub-grantees, contractors, cost-sharing partners, etc.) commit to and work toward the annual and/or long-term goals of the program?	**Condition 4** PB is used by the agency's principal to control the agency
3.1 Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?	
2.7 Are Budget requests explicitly tied to accomplishment of the annual and long-term performance goals, and are the resource needs presented in a complete and transparent manner in the program's budget?	**Condition 5** PB is used in addition to traditional budget preparation
3.2 Are Federal managers and program partners (grantees, subgrantees, contractors, cost-sharing	

	partners, etc.) held accountable for cost, schedule and performance results?	

The following five policy areas were examined which are believed to contain a suitable case in the Netherlands:

- Justice - detention
- Aviation – traffic air control
- Plant disease control
- Forestry service
- Food safety

Despite the fact that the PART analysis takes federal programs as its unit of analysis, the questions take into account a number of aspects of the principal agent relationship. The relevant principal-agent relationship is also stated in the PART analysis itself. It is therefore possible to select relevant principal-agent relationships based on these program reviews. Within this relationship, several programs may be managed simultaneously.

A condition is considered to be met if the majority of questions are answered with a YES. For condition 3 this may mean that just three out of four questions were answered with a YES while for the other two conditions, both associated questions require a yes. A suitable case is one in which at least two of conditions 3-5 are met and none on the conditions 3-5 is clearly not met. It should be noted that this analysis for conditions 3-5 and the assumption of meeting conditions 1 and 2, is a crude method of case selection. It does however give an indication in which direction to look for further analysis.

The analysis of the PART database revealed that in all five policy areas, programs existed that met the criteria. With regard to the activities of the Netherlands Forest Service it should be noted that there are four federal U.S. counterparts to choose from (National Park Service, Forest Service, Fish and Wildlife Service and the Bureau of Land Management). In addition to analysis of U.S. studies on performance information use and a discussion with Professor and former GAO director Paul Posner, it was decided to approach the FAA and the USFS as U.S. candidates. Although, based on the U.S. analysis, plant disease control and food safety appeared slightly superior these candidates were abandoned on pragmatic reasons. Mastering the essence of the specific and technical aspects of plant disease control was considered to require a large amount of time and therefore represent a risk compared to the other areas. With regard to food safety, the agency in the Netherlands was at the time engaged in a problematic merger with other organizations that also entailed the moving of headquarters to another city. Therefore it was feared that getting the right people for interviews would pose a challenge.

APPENDIX V   FORMAT FOR CASE DESCRIPTION

Paragraph 1   Description of the Agency and its Principal
- A description of organization characteristics such as age, size and organizational structure with additional focus on the relevant units.
- Policy field and tasks: the policy area, mission, objectives, tasks and activities of the agency and its agent (or relevant unit).
- Principal-agent relationship: A description of the formal relationship between the agency and its agent in formal terms as well as dominant perceptions that may exist of one another.
- A description of the actors in the budgetary process for both agency and the principal and of
  factors that dominate the allocation decisions in the policy field
- Reforms that have been made in the budgetary process, especially PB reforms

Paragraph 2   Degree of PB implementation
- An analysis of the formal PB system under which the agency operates and compliance with its requirements. This includes the link between performance and funding.
- Performance assessment by to principal: The degree to which the principal makes use of performance information reported by the agency.
- Performance management by the agency: the use of PI for program learning or enlightenment and what this contributes to improving efficiency or effectiveness.

Paragraph 3 should provide an answer with regard to these indicators:

Indicator		Presence
PB 1.1	PB is used by the agency in addition to traditional budgeting	Absent/ present to some extent / clearly present
PB 1.2	PB is used by the principal to control the agency	Absent/ present to some extent / clearly present
PB 2.1	A high degree of de jure PB implementation	Absent/ present to some extent / clearly present
PB 2.2	A high degree of de facto PB implementation	Absent/ present to some extent / clearly present

Paragraph 3   Exploring Neo-Institutionalist Explanations
- Path dependency: an analysis of important changes within the organization or its policy environment that may explain PB implementation.
- Asymmetry of power: an analysis of possible powerful champions of PB and an assessment of the relative importance of policy specialists in the policy field in question.

272

- Cultural appropriateness: an analysis of the archetype characteristics of a learning organization, applied to the agency and its principal
- Cognitive frames: An assessment of the fit between the PB indicators and the perception of success that organization members hold.

Paragraph 4 should provide an answer with regard to these indicators:

Indicator		Presence
H 1.1	Critical juncture in the accountability chain or policy field	Absent/ present to some extent / clearly present
H 1.2	Problem to which PB was seen as a solution	Absent/ present to some extent / clearly present
H 2.1	An advocate or champion of PB in a powerful position.	Absent/ present to some extent / clearly present
H.2.2	A policy field in which specialists are dominant.	Absent/ present to some extent / clearly present
S 1.1	Absence learning disabilities	Absent/ present to some extent / clearly present
S 1.2	Participative openness	Absent/ present to some extent / clearly present
S 1.3	Reflective openness	Absent/ present to some extent / clearly present
S 2.1	Shared view of the meaning of measured performance	Absent/ present to some extent / clearly present

Paragraph 4   Contextual factors
- What macro-economic factors explain the budgetary or policy decisions?
- What political preferences explain the budgetary or policy decisions?
- What other likely factors may explain budgetary or policy decisions?

Paragraph 6   Conclusions
- Assessment of the overall score on the indicators
- Which one of the formulated hypotheses does this supports
- Other relevant observations from the case regarding the central question

Epilogue
- Additional observations
- Data collection

EARLIER PUBLICATIONS AND CURRICULUM VITAE AUTHOR

Earlier publications and research papers (professional and academic):

De Jong, M. (2015), *The Impact of Performance Management on Performance in Public Organizations: Negligible yet Increasing?* Published in Public Administration Review, Volume 75, Nr.5, October 2015. Commentary on article by Dr. Ed Gerrish.

De Jong, M. (forthcoming 2015). Case studies on PBB in Poland and the Netherlands as part of a World Bank/PEMPAL publication on good and bad practices of PBB implementation in the US, France, Australia, the Netherlands, Estonia, Russia and Poland. Eds: .J.I. Beazley and D. P. Moynihan

De Jong, M. (2014), *Purposeful use of performance information in forestry and air traffic control: a matter of nature or nurture? – A transatlantic comparison.* Paper presented at La Follette School of Public Affairs at the invitation of the European Union Center of Excellence at the University of Wisconisin, October 7th, 2014 Madison. An earlier version was presented at the EGPA conference in Speyer, Germany on September 12th 2014

De Jong, M. and Van Nispen, F.K.M (2014), *On the utilization of performance information in times of austerity: from success to a return to squeaky wheel budgeting?* Paper delivered at the 26th Annual Conference of the Association for Budgeting and Financial Management (ABFM), Grand Rapids, October 1st-4th, 2014

De Jong, M., Van Beek, I en Posthumus, R (2013), *Introducing Accountable Budgeting – Lessons from a Decade of Performance Based Budgeting in the Netherlands*. Published in OECD Journal on Budgeting in July 2013 (Volume 12, Issue 3)

De Jong, M. (2013), *Recent Performance-Based Budgeting Reforms in The Netherlands – Another Lap Around the Windmill!* Posted on January 22nd 2013. For this blog I was named top author of 2013 on IMF's PFMBlog.

De Jong, M. (2013), *Using performance metrics in public sector agencies to learn and improve: a matter of institutional history and identity?* Paper presented at the 11th Public Management Research Conference, Madison, WI, June 20-22, 2013

De Jong, M. and Cammel, M.I. (2012), *Van VBTB naar VB – minder letters, meer informatie*. TPC, Tijdschrift voor public governance, audit & control, jaargang 10, nr 5, Oktober 2012, p.32-35

De Jong, M. and Van Nispen, F.K.M. (2012), *Institutional Explanations for Result Orientation: Codification or Modification?* Paper delivered at the 24th Annual Conference of the Association for Budgeting and Financial Management (ABFM), New York City, October 11th-13th, 2012

De Jong, M. and Hardt, Ł (2011), *Improving the Quality of Governance in Poland through Performance Based Budgeting.* Report from study that was awarded grant form Ernst & Young Poland's Better Government Programme. It was presented to Polish government and press in Warsaw on Dec 7th 2011.

De Jong, M. (2011), *Managing public performance through budgetary incentives: appropriate regardless of the consequences?* Paper delivered at 6th general conference of the European Consortium for Political research (ECPR), Reykjavik - August 25-27, 2011

De Jong, M. and Vos, R.O. (2011), *Vertrouwen geven en toch in control zijn – praktische handvaten voor een veelbelovend concept*. TPC, Tijdschrift voor public governance, audit & control, jaargang 9, nr. 3, Juni 2011, p.12-15

De Jong, M. and Verheij, P., (2009), '*Externe verslaglegging van de centrale overheid, hoofdstuk in 'Externe Verslaglegging van Publieke Organisaties'* red. Dees, M., (2009), Sdu Den Haag (ISBN nr 978 90 12 12875 9)

De Jong, M. (2008), Assessing the PART – *an Analysis of OMB's Program Assessment Rating Tool and Comparison to the NL Policy Review PART*, report of independent study conducted at George Mason University, Arlington, VA

De Jong, M. ((2008), *Exploring the Limits of Performance Informed Budgeting: OMB's PART versus the NL Policy Review.* Paper presented on expert-meeting under the auspices of the EGPA Study-Group on Public Sector Financial Management Vienna, Austria June 4-5, 2008

De Jong, M. (2007), *Leren van beleidsevaluatie – De beperkingen van de blik van de controller*, TPC, Tijdschrift voor public governance, audit & control, jaargang 5, nr 3, Juni 2007 p.16-19

De Jong, M. (2005), *Performance budgeting: what did it bring us so far? - a transatlantic comparison.* Abstract presented at the 17th conference of the Association for Budgeting and Financial Management (ABFM) Washington DC, in November 10-12, 2005

About the Author

Maarten de Jong (1969) grew up in Almelo in the eastern part of the Netherlands. He holds Masters in both Business Administration (Rijksuniversiteit Groningen) and Public Controlling (Erasmus Universiteit Rotterdam). Maarten has been employed in public finance for over fifteen years and has been closely involved in budget reform initiatives in different government organizations. Since 2006 he has been lecturing financial professionals on a regular basis at the educational institute of the Ministry of Finance (RAFEB) as well as occasionally to students at other institutes.

Maarten started his public finance career in 2000 as a financial advisor at the Municipal county of Hoogvliet/City of Rotterdam where he served as head the financial and quality control unit and as a member of the management team. In 2005 he moved to a central government position at the financial economic directorate of the Ministry of Infrastructure (V&W). From 2007 Maarten has been employed as a senior budget specialist at the Netherlands Ministry of Finance. As such he was responsible for a major reform of the Netherlands' performance budgeting system in 2011-2013.

Internationally Maarten regularly represents the Netherlands in international fora and communities of practice on performance budgeting and performance management. In addition he has also been involved in technical assistance to budget reforms in a number of nations in Eastern Europe, Asia and North Africa on behalf of the Netherlands government. Apart from his regular work, Maarten has been contracted for lecturing, consulting and research assignments by amongst others the OECD, the World Bank and the EC. In 2010 he was awarded a scholarship from Ernst and Young's Better Government Programme for a study on overcoming problems in Poland's performance budgeting approach. In 2013 he was named top author of IMF's Public Financial Management Blog for a contribution on performance budgeting reform in the Netherlands.

Maarten is married (2001) and has one daughter (2008). Other family members include a dog, a cat and a number of turtles and tortoises. In addition to being passionate about his family and public finance, Maarten is also an aviation enthusiast and an amateur herpetologist (= studying reptiles and ambhibians).

Contact info

m.jong@minfin.nl (work)
maarten.dejong@fsw.eur.nl (university)
maarten@dejongengeuze.com (home)