



The use of performance measurement systems in the public sector: Effects on performance



Roland F. Spek^{lé} ^{a,*}, Frank H.M. Verbeeten ^b

^a Nyenrode Business University, The Netherlands

^b University of Amsterdam, The Netherlands

ARTICLE INFO

Keywords:

Performance measurement systems
Performance contracting
Public sector
New public management
Contractibility

ABSTRACT

We study the use of performance measurement systems in the public sector. We hypothesize that the way in which these systems are being used affects organizational performance, and that these performance effects depend on contractibility. Contractibility encompasses clarity of goals, the ability to select undistorted performance metrics, and the degree to which managers know and control the transformation process. We expect that public sector organizations that use their performance measurement systems in ways that match the characteristics of their activities outperform those that fail to achieve such fit. We test our hypotheses using survey data from 101 public sector organizations. Our findings indicate that contractibility moderates the relationship between the incentive-oriented use of the performance measurement system and performance. Using the performance measurement system for incentive purposes negatively influences organizational performance, but this effect is less severe when contractibility is high. We also find that an exploratory use of the performance measurement system tends to enhance performance; this positive effect is independent of the level of contractibility. The effectiveness of the introduction of performance measurement systems in public sector organizations thus depends both on contractibility and on how the system is being used by managers. These findings have important implications, both for practice and for public policy.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past two decades, the introduction of performance measurement has been one of the most widespread international trends in public management (Pollitt, 2006). Reforms in the spirit of the New Public Management (NPM) movement have led to major changes in the management of public sector organizations, based on the notion of competitive markets and the adoption of private sector management techniques (Groot and Budding, 2008; Hood, 1995; Pollitt, 2002; Walker et al., 2011). Although specific manifestations of NPM reforms vary over time and across

countries (Kelman and Friedman, 2009; Pollitt, 2006), they share a number of common themes (Hood, 1995; Pollitt, 2002; Ter Bogt, 2008). Two common characteristics of NPM thinking are the introduction of economic rationality and efficiency as overarching principles, and its belief in the beneficial effects of business-like management practices and instruments (Ter Bogt et al., 2010), including performance measurement practices (Brignall and Modell, 2000; Broadbent and Laughlin, 1998; Groot and Budding, 2008; Hood, 1995; Pollitt, 2002, 2006).

To promote an effective, efficient, and accountable public sector, NPM advocates a rather mechanistic notion of performance contracting in which explicit and measurable pre-set performance targets should guide civil servants' efforts towards the achievement of their organizations' objectives (Bevan and Hood, 2006; Newberry and Pallot, 2004). In this cybernetic view of contracting, performance

* Corresponding author at: Nyenrode Business University, P.O. Box 130, 3620 AC Breukelen, The Netherlands. Tel.: +31 346 291225.

E-mail address: r.spekle@nyenrode.nl (R.F. Spek^{lé}).

measurement systems are supposed to create incentives that help to align individual goals with the objectives of the organization, provide valuable feedback information on the progress towards these objectives, and form the basis for internal and external accountability (Cavalluzzo and Ittner, 2004; Heinrich, 2002).

The second problem with NPM's focus on incentive-oriented performance contracting is that the assumed benefits meet uneasily with a large body of theoretical work in the academic literature. In this literature, the consensus is that incentive-oriented performance contracting of the NPM type can only work in conditions of high contractibility, i.e. when: (1) the goals of the organization are clear and unambiguous; (2) performance can be measured in ways that are consistent with organizational goal achievement; and (3) organizational actors know and control the transformation processes and are able to predict the likely outcomes of various alternative courses of action (e.g. Baker, 2002; Feltham and Xie, 1994; Gibbons, 1998; Hofstede, 1981; Otley and Berry, 1980). If these three cumulative conditions cannot be met, performance measures will only provide a partial representation of the organization's ultimate objectives, in which case a strong emphasis on result targets is likely to have dysfunctional consequences because the incentives induce organizational actors to focus on target achievement rather than on organizational goals. In a public sector context, this risk is particularly real (cf. Kelman and Friedman, 2009) because the goals of many public sector organizations are notoriously ambiguous (Burgess and Ratto, 2003; Dixit, 1997, 2002; Tirole, 1994), and the selection of appropriate performance measures is known to be difficult in this sector (Hyndman and Eden, 2000).

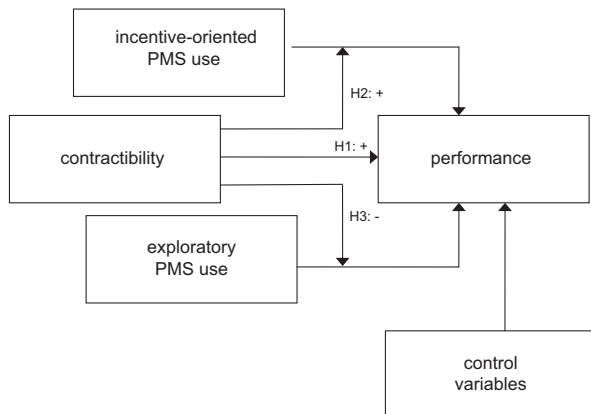


Fig. 1. Theoretical Model.

Our study aims to contribute to this debate by providing larger sample quantitative evidence on the organizational factors that moderate the effectiveness of the use of performance measurement systems in public sector organizations. We focus on two roles of the performance measurement system (incentive-oriented use and exploratory use), and on one key organizational characteristic (contractibility of performance). Building on economic and behavioural theories, we develop and test the model that is reflected in Fig. 1.

2. Theory development

2.1. Institutional characteristics of the public sector in the Netherlands

The Dutch public sector provides the setting for this empirical study. NPM has gained a strong foothold in the Netherlands, and has had considerable influence on successive government and public sector reform initiatives (Groot and Budding, 2008; Ter Bogt, 2008; Ter Bogt and Van Helden, 2005; Van Helden, 1998). A large majority of governmental organizations at the local, regional (i.e., provincial) and central government level have introduced performance measurement systems as part of the reform initiatives in the 1980s and 1990s. In addition, performance measurement techniques have been introduced in other public sector organizations such as the police, hospitals, education and research institutes, public transportation, courts, and the public prosecution service (De Bruijn and Van Helden, 2006). In this section, we discuss some of the specific institutional characteristics of this sector¹ to provide some background to our study.

The Dutch governmental system is organized in three different layers: central government, provinces and municipalities. The government system can be characterized as a decentralized unitary state. On the one hand, this means that the relations between the layers are hierarchical: municipalities have to adhere to the policies of the central government and the province. On the other hand, municipalities and provinces have considerable freedom in implementing (financial) policies. Central government is organized in Ministries (responsible for policy design) and executive agencies (responsible for policy implementation). Funding for the first group is mostly based on fixed budgets, while funding for executive agencies is largely based on production parameters. To control costs at the central government level, there is generally a 'macro-budget boundary' to make sure that costs do not exceed the allocated budget.

2.2. Contractibility

NPM advocates a rather mechanistic notion of performance contracting in which explicit and measurable performance targets should guide public sector employees efforts (cf. Newberry and Pallot, 2004). Both the behavioural and economics literature argue that such performance contracting is suitable only in specific situations. For example, Ouchi (1977) suggests that performance contracting is suitable only when knowledge about the

transformation process is available, and when outputs can be measured timely and adequately. Hofstede (1981) extends the contingency framework to public sector organizations and suggests that result controls are most useful when objectives are unambiguous, outputs are measurable, activities are repetitive and the effects of management interventions are known. The economics literature holds that the principal must be able adequately to specify and measure the results to be obtained in order to rely on performance contracting (Burgess and Ratto, 2003; Propper and Wilson, 2003; Tirole, 1994). Overall, the general theoretical consensus is that explicit performance contracting requires (1) that goals can be specified unambiguously in advance; (2) that the organization is able to select undistorted performance measures, i.e.

H1. Contractibility is positively associated with performance.

2.3. *The use of performance measurement systems*

Performance measurement systems may serve a variety of different purposes within organizations. Fig. 2 summarizes the taxonomy that we use in this paper, and sketches its relation to the other classifications discussed in this section.

Empirical studies addressing the different roles of these systems have relied on several classifications to define and operationalize performance measurement system use.

For the purposes of our paper, we draw on a strand of literature that refers to the ‘dual role of controls’ (cf. Tessier and Otley, 2012), i.e., the classical distinction between the decision-facilitating and decision-influencing roles of performance measurement systems (cf. Ahrens and Chapman, 2004; Demski and Feltham, 1976; Sprinkle, 2003). The decision-influencing role refers to the use of information for motivating and controlling managers and employees (‘incentive-oriented use’), while the decision-facilitating role refers to the provision of information to guide decisions and managerial action (cf. Grafton et al., 2010; Van Veen-Dirks, 2010). With regard to the decision-facilitating role, we follow Hansen and Van der Stede (2004) and distinguish between single-loop learning (‘operational use’) and double-loop learning (‘exploratory use’).² Each of these roles will be discussed briefly.

2.3.1. *Incentive-oriented use*

As we seek to examine the effects of NPM-consistent ways of using the performance measurement system, one of the roles in the taxonomy should match NPM’s views on the contribution of performance measurement to the functioning of public sector organizations. The NPM programme emphasizes the role of the performance measurement system in target setting, incentive provision, and rewards (Newberry and Pallot, 2004). We refer to this role as an incentive-oriented use of the performance measurement system.

2.3.2. *Operational use*

Operational use is similar to the operational planning (Hansen and Van der Stede, 2004), progress monitoring (Franco-Santos et al., 2007) and monitoring (Henri, 2006) role; it involves operational planning, process monitoring and provision of information. Previous research (e.g. Hansen and Van der Stede, 2004) indicates that this role is prevalent across organizations, suggesting that it is a basic requirement rather than the result of a deliberate organization-specific or situation-dependent design choice. Therefore, this role is not the main focus of our research.

2.4. *Contractibility and performance measurement system use*

As mentioned previously, the general theoretical consensus is that explicit performance contracting requires high contractibility, i.e., (1) unambiguous goals (2) undistorted performance measures, and (3) knowledge and control of the production function (Baker, 2002; Feltham and Xie, 1994; Gibbons, 1998; Hofstede, 1981; Otley and Berry, 1980). The first requirement is self-evident: if the organization is unable to specify what it wants from its managers, performance metrics cannot provide much guidance, and management control becomes political control (Hofstede, 1981). If goals are complex and ambiguous, performance metrics can at best provide a partial representation of the organization’s ultimate objectives. As performance based contracts induce agents to engage in behaviours that positively affect their score on the measures, incomplete metrics lead managers to pay unbalanced attention to results that are being measured, while unduly neglecting areas for which performance is not assessed (Dixit, 2002; Prendergast,

H2. The performance effect of using the performance measurement system for incentive purposes is more positive for high contractibility activities than for low contractibility activities.

In conditions of low contractibility, the role of performance measures is (or should be) different. For low contractibility activities, performance measures can no longer serve as a basis for performance contracting yet these measures can still be useful to communicate priorities and expectations as to desired performance (cf. Hall, 2008), as well as promote creativity and flexibility to achieve the communicated goals (cf. Mundy, 2010). In such situations, the emphasis should not be on meeting pre-set targets, but on experimentation, double loop learning, mutual consultation, and sequential adaptation to emerging insights, gradually leading to a shared set of goals and a sense of how these goals can be achieved. An exploratory use of the performance measurement system accommodates this.

Table 1
Respondents and their organizations.

	Mean	Std. dev.	Min.	Median	Max.
Experience in organization (years)	14.5	10.1	2.0	11.5	39.5
Experience in current position (years)	5.0	4.7	1.0	4.0	34.5
Size organization (FTE)	1983	3556	26	950	30,000
Size unit (FTE)	206	400	5	65	3000

Table 2
Composition and reliability of performance and contractibility variables.

Variable	Items	Component loading
PERFORM (unit performance) Cronbach's $\alpha = 0.770$	• productivity	0.701
	• quality or accuracy of work produced	0.673
	• number of innovations, process improvements, or new ideas	0.547
	• reputation for work excellence	0.711
	• attainment of production or service level goals	0.750
	• efficiency of operations	0.592
	• morale of unit personnel	0.580
CLEARGOALS (clarity of mission and goals; formative indicator of contractibility) Cronbach's $\alpha = 0.768$	• univocality of mission statement	0.863
	• mission is written on paper and is communicated internally and externally	0.854
	• unit goals are unambiguously related to the mission	0.768
	• unit goals are specific and detailed	0.552
MEASGOALS (measurability of output in goal consistent way; formative indicator of contractibility) Cronbach's $\alpha = 0.717$	• the set of performance metrics provides a complete picture of the results to be achieved	0.677
	• the performance measures of the unit are unambiguously related to the goals of the organization	0.722
	• goal achievement depends heavily on external factors (reversed scored)	0.572
	• the causal relation between resource allocation and goal achievement is clear	0.701
	• the effect of our efforts become visible within a year	0.701
TRANSFORM (knowledge of transformation processes; formative indicator of contractibility) Cronbach's $\alpha = 0.714$	• in performing our tasks, there is a logical way to proceed	0.620
	• the unit's primary processes can only be performed in one specific and documented way	0.678
	• within the unit, it is entirely clear how to perform our tasks	0.812
	• in performing their tasks, employees rely on standard procedures and rules	0.832

ment described and analyzed in [Whitney et al. \(1983\)](#). Based on factor analysis results (see [Table 2](#)), we compute CLEARGOALS, MEASGOALS, and TRANSFORM by averaging the relevant item scores. The alphas of the resulting variables range from 0.714 to 0.768, indicating adequate reliability. We then compute our contractibility proxy CONTRACT by summing the scores on the formative indicators CLEARGOALS, MEASGOALS, and TRANSFORM.⁷

Table 3
Performance measurement system use: component loadings and reliability.

performance measure type ($a = \text{input}$; $b = \text{process}$; $c = \text{output}$; $d = \text{quality}$; $e = \text{outcome}$)		a	b	c	d	e
OP-USE (operational use) Cronbach's $\alpha = 0.877$	• operational planning		0.648	0.545	0.735	0.717
	• budget allocation		0.606	0.650	0.685	0.604
	• process monitoring		0.711	0.525	0.609	0.976
INC-USE (incentive purposes) Cronbach's $\alpha = 0.905$	• career decisions	0.596	0.738	0.834	0.676	0.721
	• bonus determination	0.624	0.823	0.803	0.796	0.752
EXPL-USE (exploratory use) Cronbach's $\alpha = 0.925$	• communicating goals and priorities	0.725	0.707	0.638	0.563	0.738
	• evaluating appropriateness of goals and policy assumptions	0.649	0.720	0.745	0.653	0.758
	• policy revision	0.646	0.796	0.759	0.688	0.742

upon one factor. The CFA (presented in [Table 3](#)) supports our expectations,⁸ leading to three multi-item variables representing performance measurement system use. For example, operational use component loadings for process measures (performance measure type b in [Table 3](#)) are 0.648 (operational planning), 0.606 (budget allocation) and 0.711 (process monitoring), respectively; operational use component loadings for output measures (performance measure type c) are 0.545 (operational planning), 0.650 (budget allocation) and 0.525 (process monitoring). For all three variables, Cronbach's alpha values exceed 0.85, suggesting adequate reliability. We calculate OP-USE, INC-USE and EXPL-USE by taking the mean of the item scores that have component loadings above 0.5 on the identified use-variable. Note that our analysis does not exclude the possibility that the resulting scales for performance measurement use are closely related; we deal with this issue in subsequent sections of the paper.

4. Results

4.1. Descriptive statistics

[Table 4](#), panel A presents summary statistics for each variable. [Table 4](#), panel B gives the correlation matrix. From this table, we observe that the various uses overlap considerably: the Pearson correlation between operational use (OP-USE) and exploratory use (EXPL-USE) is 0.741, while the correlations between OP-USE and incentive-oriented use (INC-USE) and between INC-USE and EXPL-USE are 0.521 and 0.527 respectively. Apparently, the three roles

of the performance measurement system are not mutually exclusive, and these bivariate correlations indicate that if an organization intensifies its use of the performance measurement system for a specific purpose, it is likely to intensify its use for the other purposes as well.

4.2. Modelling issues

We test our hypotheses by estimating a multivariate OLS model. In the analysis, we use White heteroskedasticity-consistent standard errors and covariance to correct for possible heteroskedasticity. The model features two multiplicative terms to capture the hypothesized moderator effect of contractibility on the performance effects of both an incentive-oriented and an exploratory use of the performance measurement system. Our theory does not predict any direct effects of performance measurement system use on performance. The model nevertheless tests for the main effect of the relevant USE variables to ensure that significant coefficients for the interaction effects are not in fact due to lower order effects (cf. [Echambadi et al., 2006](#); [Hartmann and Moers, 1999](#)). Furthermore, we include operational use (OP-USE). We have no specific hypotheses on operational use, but wish to explore its role empirically as most organizations appear to use performance measures for operational purposes (cf. [Hansen and Van der Stede, 2004](#)). Finally, the model incorporates organizational and unit size, and dummies for central government and local government to control for potential size effects and for possible branch-related influences.

5. Conclusions and discussion

This study is one of the first larger scale empirical studies to provide evidence on the effects of the use of performance measurement systems in public sector organizations. Our results provide some interesting insights into the functioning of these organizations. First, we find a positive association between contractibility and performance. This finding is consistent with a large literature documenting the positive performance effects of clear and measurable goals ([Locke and Latham, 2002](#)).

Our results also indicate that the current NPM-approach to performance contracting as a means to improve public sector performance may only hold for a subset of public sector organizations, i.e., those that are characterized by high contractibility activities. More specifically, we find that an incentive-oriented use of the system negatively influences performance, but that this effect becomes less negative if contractibility increases. These findings imply that NPM cannot maintain its universalistic pretensions, and should allow for a more situation-dependent approach to performance management. One might even be tempted to conclude that NPM is ill-founded generically, considering the strong direct negative effect of an incentive-oriented use on performance. This conclusion, however, would overstretch the evidence. In our sample, we observe that the emphasis on incentives is generally quite low.¹² Thus, the sample contains hardly any units that pair high contractibility with high incentive use. This is typical for the Dutch public sector, which is characterized by a low emphasis on incentives. It is conceivable that if 'high contractibility, high incentive' units were part of the sample (quod non), we would have found a stronger performance effect for incentive-oriented use in case of high contractibility. Or to rephrase the issue: it is possible that the high contractibility units in our sample would have done better had they increased their emphasis on incentives. But this is speculative, and our data provide no indication that this is in fact the case.

We also found that the exploratory use of performance measures enhances performance. Apparently, performance measures provide public sector managers with the opportunity to communicate goals more clearly, to evaluate the appropriateness of the goals and policy assumptions, and to revise policies. While such exploratory use is not explicitly recognized by NPM, it may be one of its unintended consequences (cf. [Marsden and Belfield, 2006](#)). Interestingly, the positive performance effect of this exploratory use appears to exist independent of the level of contractibility. This finding contradicts our hypothesis; we expected that performance would increase only if contractibility is low. As before, this finding may also

be due to the general reluctance to an incentive-oriented use of performance measures we observe in our data. In the high contractibility condition, an exploratory use might be a second-best solution; even if an incentive-oriented use is best, using the system in an exploratory way may be preferable to not using the system at all. If, however, no one opts for the first-best solution, the relative inferiority of the exploratory use will not show up in the analysis. Another explanation may be in the fact that the exploratory use of performance measures communicates strategy, and that this communication eliminates ambiguity and confusion about objectives.

Acknowledgements

We are grateful to our students in the Nyenrode Public Sector MBA programme for their invaluable help in collecting the data. This paper has benefited from insightful comments and suggestions from three anonymous reviewers. Earlier versions of this paper have been presented at the 5th International Conference on Accounting, Auditing & Management in Public Sector Reforms, the EIASM New Directions in Management Accounting Conference, the AAA MAS Conference, and a workshop session at Tilburg University. We thank conference participants for their feedback, especially Frank Moers and Naomi Soderstrom. We also appreciate helpful discussions with Sally Widener and Anne-Marie Kruis.

Note: this question is repeated five times, i.e. for each type of measure (input, process, output, etc.) separately. The categories to which we refer in the question are part of the common vocabulary in the public sector, and we assumed that our respondents knew these terms. This assumption was in fact corroborated in the pre-test of the questionnaire. Nevertheless, to further ensure a sufficient understanding, we include per type of measure a question asking respondents to indicate whether or not they use that specific type of measure, providing a number of representative examples. These questions immediately preceded the questions on the way in which these measures were being used. Items d, j and k have been dropped from the analysis because of an insufficiently clear conceptual link with the theoretical constructs we address in the analysis.