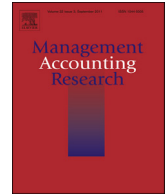




Contents lists available at ScienceDirect

Management Accounting Research

journal homepage: www.elsevier.com/locate/mar



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

Roland F. Spek^l, 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^l).

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).

NPM's emphasis on the role of performance measurement systems in target setting, performance evaluation, and incentive provision is, however, problematic for two main reasons. The first of these is that NPM's exclusive focus on the role of performance measurement systems in accountability and incentive provision is too narrow. In the academic literature, it is widely acknowledged that performance measurement systems can serve various different purposes, and that they can be applied in a variety of different ways (e.g. Franco-Santos et al., 2007; Hansen and Van der Stede, 2004; Henri, 2006; Simons, 1990). By ignoring these other roles, NPM disregards the performance consequences of alternative ways of using the performance measurement system.

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).

This problem is not just hypothetical. While analytical (Dewatripont et al., 1999) and practitioner (Wilson, 1989) literature emphasizes the benefits of the NPM-approach, several small-sample studies provide qualitative evidence of dysfunctional effects of performance contracting in the public sector (e.g. Bevan and Hood, 2006; De Bruijn, 2002; Newberry and Pallot, 2004). But larger scale empirical evidence on the organizational conditions that are appropriate for applying performance contracting in the public sector is still remarkably scarce (Goddard, 2010; Hood, 2007; Taylor, 2011; Van Helden, 2005), and at present, the question as to the effect of NPM-type incentive-oriented contracting on organizational performance cannot be answered conclusively.

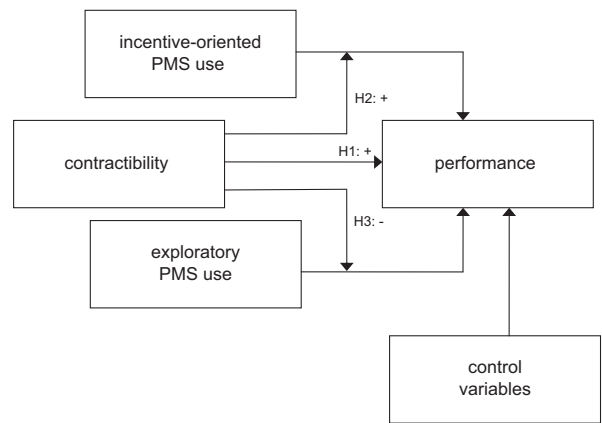


Fig. 1. Theoretical Model.

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.

We will argue in subsequent sections of this paper that the cybernetic, incentive-oriented use of performance measures as advocated by NPM enhances performance only when contractibility is high. Relying on an incentive-oriented use when contractibility is low is likely to result in an unbalanced attention to quantifiable results, 'managing measures rather than results', and data manipulation (Marsden and Belfield, 2006; Newberry and Pallot, 2004). When contractibility is low, performance may benefit from an exploratory use of performance measures as this provides opportunities for double loop learning, mutual consultation, and adaptation to emerging insights (cf. Burchell et al., 1980). Our central claim is that public sector organizations that use their performance measurement systems in ways that are consistent with the characteristics of their activities outperform organizations that fail to achieve such fit.

We empirically test the model using survey data from 101 public sector organizations. The results partially confirm our theoretical expectations. Specifically, we find that contractibility moderates the relationship between an incentive-oriented use of the performance measurement system and performance: while using the performance measurement system for incentive purposes negatively influences organizational performance, this effect is less severe when contractibility is high. This finding suggests that NPM should allow for a more situation-dependent approach to performance measurement. We also find that an exploratory use of the performance measurement system tends to enhance performance in public sector organizations. Contrary to our expectations, this positive effect appears to exist independent of the level of contractibility. These results suggest that public sector managers need not

only consider the technical aspects of performance measurement system design, but also how they are going to use the system. Jointly, the evidence indicates that both the alignment with the activities of the organization and the way in which public sector managers use their performance measurement system significantly affect organizational performance in the public sector.

The remainder of this paper is structured as follows. In Section 2, we provide insight into the institutional characteristics of the Dutch public sector, review the relevant literature and develop our theory and hypotheses. In Section 3, we provide details on our methods. We report on the empirical analysis in Section 4. Section 5 presents conclusions and a discussion of the findings.

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.

Municipalities and provinces obtain their funds partially from central government. Some of these funds are

earmarked for specific goals, yet most of the funds are generic. Democratically elected councils of the municipalities and provinces decide upon the allocation of these generic funds. Central funding of lower government levels is mostly based on general characteristics (number of inhabitants, road kilometres, etc.) and is independent of performance. There is no link between efficiency and funding, and less efficient organizations need to find funding for their deficits elsewhere in their budget, or need to cut back on service levels. In addition to the central funds, municipalities and provinces obtain their resources from local taxes and revenues from the provision of goods and services (passports, building permits, etc.). Decentralization of decisions to lower government levels also encompasses organizational matters, and municipalities and provinces have considerable freedom in the design and use of their performance measurement systems.

Similar structures apply to other public sector organizations. For example, funding in health care and education is typically based on fixed fees as well as certain output parameters (for example, volume of services provided, number of graduated students, etc.). Organizations in these branches need to report quite extensively on their processes and output, but they are not required to use this information for internal management purposes. Much of the output information is publicly available, but is not systematically used for benchmarking or to construct 'league tables'. Although health care organizations and schools are subjected to centralized supervision (i.e. the health or education inspection), these supervisory bodies focus on maintaining some minimum level of service quality rather than on efficiency or effectiveness. The supervisory bodies have limited intervention rights, and can only intervene in case of severe and prolonged malfunctioning.

Summarizing, most public sector organizations in the Netherlands have considerable freedom with regard to the performance measurement practices they implement. Funding is partially based on size criteria and production parameters; however, the financial consequences of (not) meeting performance targets are relatively small. While in some sectors the available performance information could be used for ranking purposes (e.g. in health care and education), the impact of these rankings at this moment is still rather limited. Most of the performance measurement information appears to be used as 'intelligence systems' (Hood, 2007) to provide background information. Only when organizations are on the brink of bankruptcy, higher government layers become more closely involved in managing the organization.

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

¹ We rely on Groot and Van Helden (2012), Ter Bogt and Van Helden (2005), and Pollitt and Bouckaert (2011).

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. metrics that provide incentives that are adequately aligned with the organization's ultimate objectives; and (3) that organizational actors know and control the production function that transforms efforts into results, and are able to predict the likely outcomes of alternative courses of action (Baker, 2002; Feltham and Xie, 1994; Gibbons, 1998; Hofstede, 1981; Otley and Berry, 1980). In this paper, we use the term 'contractibility' to refer to the degree in which these three cumulative conditions can be met simultaneously. Examples of highly contractible activities in the public sector include garbage collection or the exploitation of a public swimming pool. In contrast, low contractibility is present when actors are unable fully to specify the attributes of satisfactory performance, or when the manager's systematic influence on the ultimate outcome is restricted or unknown. In a public sector context, this would for instance be the case in child protection services, in foreign affairs, or in the activities of a community development worker.

In addition to the moderating effects of contractibility on the use of performance measures (which we discuss in subsequent sections), we hypothesize that contractibility has a direct effect on performance. There is a large body of literature that shows that clear and measurable goals contribute to performance, and this effect is stronger for relatively simple tasks (see Locke and Latham, 2002, for an overview); this refers to our high contractibility concept. Although much of this literature is based on experiments and may be criticized for its limited external validity, various archival and survey-based studies report similar effects (see Chun and Rainey, 2005; Jung and Rainey, 2008, for empirical evidence from the public sector). Therefore:

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 instance, Abernethy and Brownell (1999) and Naranjo-Gil and Hartmann (2007) follow Simons (1990) in his broad distinction between diagnostic and interactive uses. Hansen and Van der Stede (2004) identify four different roles: operational planning, performance evaluation, communication of goals, and strategy formation. Henri (2006) also proposes four types of performance measurement system use, but his classification features monitoring, attention focusing, strategic decision making, and legitimization. Franco-Santos et al. (2007) provide yet another classification. In an extensive review of the performance measurement literature, they identify 16 different roles performance measurement systems may play in organizations. They group these roles into five broad categories: (1) measuring performance, including progress monitoring, measuring and evaluating performance; (2) strategy management, which encompasses planning, strategy formulation/implementation/execution, attention focusing, and alignment; (3) internal and external communication, benchmarking, and compliance with regulations; (4) influencing behaviour, comprising rewarding behaviour, managing relationships, and control; and (5) learning and improvement, which captures feedback, double loop learning, and performance improvement (Franco-Santos et al., 2007). The main difference between the various classifications of performance measurement system uses is located in the number of roles they identify, and in the boundaries between these roles.

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 We deviate from Hansen and Van der Stede (2004) by considering the communication of goals and the formation of strategy as one type of performance measure use ('exploratory use').

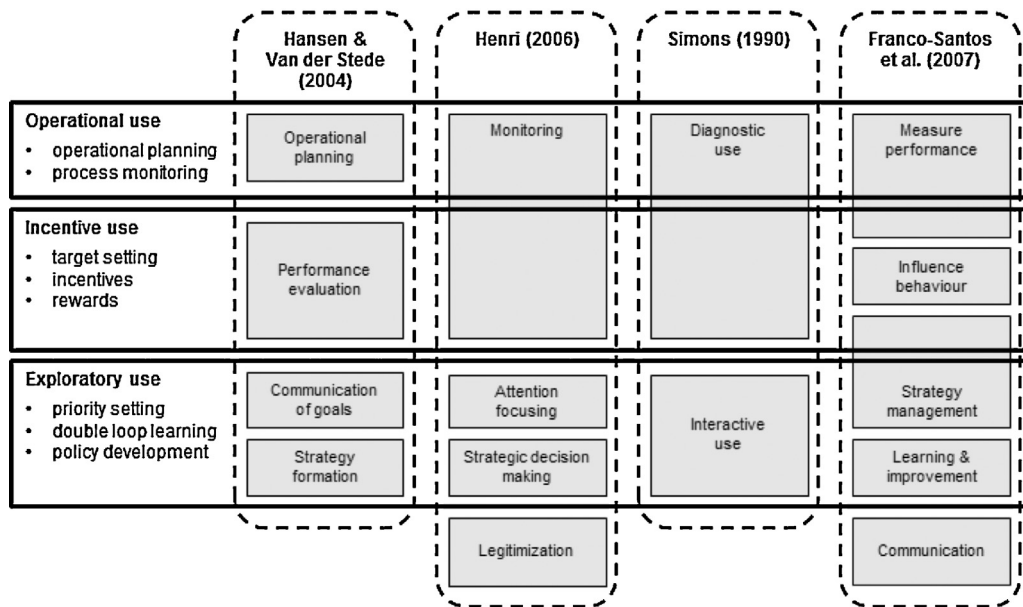


Fig. 2. Classifications of performance measurement system use.

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 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.

2.3.3. Exploratory use

An exploratory use of the performance measurement system (cf. Speklé, 2001) is conceptually associated with Simons' (1990) interactive use, and corresponds to strategy formation and communication of goals (Hansen and Van der Stede, 2004), strategy management and learning (Franco-Santos et al., 2007), and the strategic decision making role (Henri, 2006). The exploratory use of performance measurement system may help to identify policy areas that are in particular need of political or managerial attention, facilitates priority setting and selective intervention, and enables a purposive search for new policy approaches. Such exploratory use gradually leads to a shared frame of reference as to what constitutes satisfactory performance, and provides input on how this can be achieved (cf. Burchell et al., 1980; Mundy, 2010; Simons, 1990; Speklé, 2001). Rather than focusing on keeping things on track, this performance measure use is about the discovery of sensible tracks.

Next, we discuss how contractibility theoretically moderates the impact of the incentive use and exploratory use on the effectiveness of public sector organizations.

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, 1999).

The second requirement indicates that performance metrics should be 'goal-congruent' (Feltham and Xie, 1994) and 'undistorted' (Baker, 2002). Distortion is present when managerial action choices that contribute to organizational goal achievement are not (or not fully) picked up by the performance measurement system. It is also present when managers can choose actions that make them look good on the performance measures, without actually enhancing goal achievement. In both situations, the performance measurement system provides incentives that stimulate managerial behaviour that is not consistent with the ultimate objectives of the organization.

The third requirement refers to the problem of controllability and noise. Result-oriented performance contracting assumes that the manager has a significant influence on the outcomes of the activities for which he or she is held accountable, and that he or she is able to work systematically towards goal achievement. In public sector organizations, means-end relations are often ill-understood, and managers may be unable to predict the likely outcomes of alternative courses of action. Furthermore, it is often the case that realized performance as it gets measured depends not only on the individual manager's efforts and action choices, but also on decisions taken elsewhere in the organizations, on the cooperation of external parties, and on uncertain, uncontrollable events. In these situations, performance measures become noisy and do not adequately reflect managerial achievements. Noisy metrics are less informative about managerial performance, and may impose undue risk on the manager if his or her compensation or career prospects are based on these performance measures (Holmstrom, 1982; Roberts, 2004). For these reasons, noise decreases the metrics' usefulness for motivational purposes.

Consistent with the line of reasoning presented above, we argue that the effects of incentive-oriented use on performance will be contingent on contractibility. In conditions of high contractibility, not using the system in an incentive-oriented way deprives the organization of a powerful means to provide direction and goal-consistent motivation, leading to efficiency losses. However, it is also problematic to use the system for incentive purposes when conditions do not allow this; an incentive-oriented use of performance measures for low contractibility activities will result in distortion of managerial effort and dysfunctional behaviour. Both types of mismatches³ will have a negative performance effect (i.e. will lower the performance of the organization). This line of reasoning is reflected in our second hypothesis:

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.

³ Analogous to statistical terminology, Hofstede (1981) refers to these mismatches as type I and type II errors, respectively.

Using the measurement system to explore new policy options and, more generally, to foster double loop learning might seem valuable not only when contractibility is low, but also in conditions of relatively high contractibility. There is, however, a trade-off. An exploratory use of the performance measurement system requires the organization to relax its performance targets, opening them up to discussion and debate. The organization needs to relieve the evaluative pressure that goes with an incentive-oriented use of the system, because that pressure intensifies the fear of failure within the organization. This fear is inimical to innovation, for it decreases employees' willingness to engage in experimentation, and leads to a premature closing of learning (cf. Argyris, 1977, 1990; Campbell et al., 2008; Lee et al., 2004; Sprinkle et al., 2008). Relieving the pressure, however, dilutes incentives, and the cost of this dilution might well outweigh the incremental value of learning. In addition, an exploratory use of the performance measurement system requires top management time, broad organizational involvement and increased interaction among members of the organization at various hierarchical levels (Simons, 2000), all of which are costly. Therefore, the exploratory use of performance measures is only likely to enhance performance when the value of search and learning is high (Abernethy and Brownell, 1999), i.e. when the additional insights acquired in the process are likely to contribute substantively to improved policy decision making and resource allocation. This is the case in conditions of low contractibility. This results in our last hypothesis:

H3. The performance effect of using the performance measurement system for exploratory purposes is more positive for low contractibility activities than for high contractibility activities.

Fig. 3 summarizes hypotheses H2 and H3.⁴

In the development of the taxonomy (see Section 2.3), we also identified a third role of the performance measurement system, i.e. its application for operational purposes. We refrain from formulating a specific hypothesis on the effect of this type of use, but we include operational use in the analysis to explore its impact empirically and to control for any effects it may have on the relationships in the model.

3. Research design, sample, and measurement

3.1. Design and sample

The analysis is based on survey data from 101 organizational units within the Dutch public sector. In building

⁴ Our framework somewhat resembles the framework by Goold and Quinn (1990) (thanks to one of the reviewers for pointing this out to us). However, Goold and Quinn (1990) indicate that strategic control systems are 'problematic' when environmental turbulence is high and strategic objectives are difficult to specify and measure. Our framework suggests that performance measures are still useful in case of low contractibility as they provide opportunities for communicating desired performance and may signal the need for reformulation of strategies.

		Incentive-oriented	Exploratory
Contractibility	Low	MISFIT lower performance due to distortion	FIT high performance
	High	FIT high performance	MISFIT lower performance due to inefficiencies

Fig. 3. Contractibility, performance measurement system use, and performance.

the questionnaire, we relied as far as possible on instruments validated in previous studies. Also, we pre-tested the survey instrument by asking five public sector managers to complete the questionnaire and to discuss their experiences with us. This has led to some minor adjustments in the original survey design. The relevant parts of the questionnaire are reproduced in the [Appendix](#).

Respondents are managers of organizational units. We define an organizational unit as a more or less unified administrative entity within the larger organization in which the unit's manager has at least some degree of authority over the set of tasks and processes of the unit (cf. [Cavalluzzo and Ittner, 2004](#)). The organizational unit is the appropriate level of analysis for our project. Especially in larger organizations, performance measurement practices within the organization and the degree of contractibility may differ widely. This intra-organizational variety, however, is less of a problem at the level of the organizational unit, where activities and performance measurement choices are generally more homogeneous. Our dataset contains only observations from units engaged in activities that are somehow typical of the public sector. Thus, we have excluded more generic support units such as human resource management or finance departments from our sample. We have also excluded small units (number of employees <5 FTE) and units with newly appointed managers (years of experience in current position <1).

To mitigate response problems and respondent identification issues, we rely on students from a public sector MBA programme to contact potential survey participants. As a result of this procedure, our sample is not random. To the best of our knowledge, however, this procedure has not introduced systematic biases in respondent selection, and we believe that the sample is at least informative of contemporary public sector practices in the Netherlands. The sample is quite broad and diverse. Our observations include data from central government (44%), local government (27%), and various other public sector organizations (e.g. the police, schools, hospitals, social housing, et cetera; 30%). Also, the units differ widely in terms of their function and size. On average, respondents have been working for 14.5 years in their organization (median: 11.5 years) and have been employed in their current function for 5 years (median: 4 years). These figures suggest that respondents are well informed about the performance measurement practices of their unit. [Table 1](#) summarizes the information on respondents, their organizations, and their units.

3.2. Variables

3.2.1. Performance

The dependent variable is unit performance (PERFORM). We measure performance with a well-established instrument developed by [Van de Ven and Ferry \(1980\)](#). This instrument is specifically designed to capture performance in public sector organizations, and has also been used by [Dunk and Lysons \(1997\)](#), [Verbeeten \(2008\)](#), and [Williams et al. \(1990\)](#). The performance dimensions include (1) productivity, (2) quality or accuracy of work produced, (3) number of innovations, process improvements, or new ideas, (4) reputation for work excellence, (5) attainment of production or service level goals, (6) efficiency of operations, and (7) morale of unit personnel. Respondents were asked to indicate the score of their unit on each dimension, using a five-point semantic scale (1 = far below average; 5 = far above average). Factor analysis shows that these seven dimensions load on a single factor, and we calculate PERFORM by taking the mean of the item scores. Cronbach's alpha is 0.770, which is satisfactory; further details are in [Table 2](#).⁵

3.2.2. Contractibility

Consistent with the conceptual definition of the construct (see [Section 2.2](#)), we model contractibility as a formative construct⁶ that is defined by three indicators, i.e. the clarity of the organization's mission and goals (CLEARGOALS), the measurability of outputs in terms of the goals (MEASGOALS), and knowledge of the transformation process (TRANSFORM). These three formative indicators, in turn, are reflective constructs, which we measure with a variety of questionnaire items designed to pick up on a broad range of potential manifestations of these constructs. The questionnaire items relating to CLEARGOALS and MEASGOALS are based on [Verbeeten \(2008\)](#). TRANSFORM is measured with a slightly adapted version of the instru-

⁵ It should be noted that the performance dimensions included in the scale relate to output rather than outcomes. It captures for instance productivity and quality of goods and services, but it does not explicitly consider the unit's contribution to the realization of broader policy goals, and may not fully reflect the societal dimension of organizational effectiveness.

⁶ For recent treatments of the differences between formative and reflective indicators, and for a discussion of the consequences of these differences for model specification and measurement, see for instance [Bisbe et al. \(2007\)](#), [Diamantopoulos and Siguaw \(2006\)](#), and [Jarvis et al. \(2003\)](#).

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
CLEARGOALS (clarity of mission and goals; formative indicator of contractibility) Cronbach's $\alpha = 0.768$	• morale of unit personnel	0.580
	• 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
MEASGOALS (measurability of output in goal consistent way; formative indicator of contractibility) Cronbach's $\alpha = 0.717$	• unit goals are specific and detailed	0.552
	• 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
TRANSFORM (knowledge of transformation processes; formative indicator of contractibility) Cronbach's $\alpha = 0.714$	• the effect of our efforts become visible within a year	0.701
	• 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 [Whithey 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

3.2.3. Performance measurement system use

To capture the way in which performance information is being used in the public sector, we ask respondents to indicate the extent to which they use various categories of performance metrics for a variety of purposes. The categories encompass input measures (e.g. budgets, expenditure limits), process measures (e.g. efficiency, capacity use), output measures (e.g. revenues, productivity), quality measures (e.g. customer satisfaction, number of complaints), and

outcome or effect measures (realization of policy goals). For each of these 5 categories, we elicit information on the purposes for which the measures in this category are being used (1 = not at all, 5 = very frequently). Based upon theoretical reasoning (see [Section 2.3](#)), we distinguish between three purposes: operational use, incentive use and exploratory use. Operational use (OP-USE) includes the use of performance metrics (input measures, process measures, output measures, quality measures and effect measures) for operational planning, budget allocation, and monitoring of organizational processes. Incentive provision (INC-USE) encompasses the importance of performance measures in career and bonus decisions. Finally, exploratory use (EXPL-USE) comprises a reliance on performance metrics in communicating goals and priorities, in policy revision, and in evaluating the appropriateness of current objectives and policy assumptions.

We use confirmatory factor analysis (CFA) to investigate whether the empirical data match our theoretical constructs. As we have 5 categories and 11 purposes (see also the [Appendix](#)) for 101 observations, a 'regular' CFA is not possible. Therefore, we test whether the performance measure types within one purpose load

⁷ Because the scores on the formative indicators do not need to covary, factor analysing the components of CONTRACT is inconsistent and meaningless, and so is the Cronbach-type analysis of reliability at this level.

Table 3

Performance measurement system use: component loadings and reliability.

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

Cells provide component loadings for factors which we label operational use (OP-USE), incentive purposes (INC-USE) and exploratory use (EXPL-USE) for each of the performance measure types that we recognize in our survey (respectively input, process, output, quality or outcome measures).

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.

3.2.4. Control variables

To control for a possible size effect, we measure the size of the organization (SIZEORG) and the unit (SIZEUNIT) by taking the log of the number of employees. We also control for potential branch-related effects by including dummy variables for central government (CG), local government (LG), and other public sector organizations (OPS).

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

⁸ Except for the use of input measures for purposes of operational planning, budget allocation, and process monitoring (cell OP-USE, performance measure type *a*, input). Contrary to our expectations, these items do not load sufficiently strongly on operational use. The reason for this is that practically all respondents report a relatively intense use of input measures for these purposes.

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. This observation might reflect differences in management style across organizations, e.g. a 'numbers-based' style versus a more intuitive approach to management. It is also possible that these correlations indicate the existence of 'natural' spill-over effects. For instance, if an organization relies relatively heavily on its performance measurement system for planning purposes, this performance information becomes more salient, which may induce a more critical attitude towards the quality of that information. This more critical attitude will be picked up in our survey as a more intense use of the system for exploratory purposes. Similarly, if some performance metric is highly relevant for operational purposes, it is likely that the same metric is also informative of organizational and managerial performance, suggesting that it will be used for incentive purposes as well. Be that as it may, the various uses are still distinct constructs, both conceptually (see Section 2.3 for their definition) and empirically.⁹ However, these correlations signal a potential multicollinearity problem. We address this potential problem in the multivariate analysis on which we report in the following sections.

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

⁹ The internal reliability of all three use-variables is larger than the bivariate correlations between these variables, which is an indicator of discriminant validity (Churchill, 1979).

Table 4

Descriptive statistics and correlations.

Panel A: descriptive statistics						
Variable	Mean	Std. dev.	Theoretical range	Minimum	Maximum	
PERFORM	3.539	0.493	1–5	2.14	4.71	
CONTRACT	10.211	1.593	3–15	6.85	13.55	
OP-USE	3.068	0.750	1–5	1.33	4.92	
INC-USE	2.249	0.829	1–5	1.00	4.50	
EXPL-USE	3.141	0.746	1–5	1.00	4.80	
Panel B: correlation matrix						
	1	2	3	4	5	6
1: PERFORM	1					
2: CONTRACT	0.342**	1				
3: OP-USE	0.187*	0.420***	1			
4: INC-USE	0.041	0.295***	0.521***	1		
5: EXPL-USE	0.345***	0.433***	0.741***	0.527***	1	
6: SIZEORG	0.008	0.067	−0.045	−0.092	−0.220**	1
7: SIZEUNIT	0.117	0.235**	0.186*	−0.047	−0.012	0.490***

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (2-tailed).

PERFORM = unit performance; CONTRACT = contractibility (clear and measurable goals, knowledge of transformation processes); OP-USE = operational use of performance measures; INC-USE = incentive-oriented use of performance measures; EXPL-USE = exploratory use of performance measures; SIZE-ORG = size organization (= log fte organization); SIZEUNIT = size unit (= log fte of the unit).

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.

Including product terms in the model may introduce multicollinearity. Following standard practice, we standardize the relevant variables prior to forming the multiplicative terms (cf. Aiken and West, 1991; Jaccard et al., 1990). This standard practice, however, is no longer uncontroversial and there is a growing literature that argues that standardizing (or mean centring) does not affect collinearity (see for instance Echambadi et al., 2006; Echambadi and Hess, 2007; Gatignon and Vosgerau, 2005). For this reason—and also because of the relatively high correlation between operational and exploratory use (see Section 4.1) – we pay special attention to the analysis of the condition indices and the variance inflation factors (VIFs). The examination of the condition indices reveals no signs of multicollinearity. The analysis of the VIF scores is also reassuring: the highest VIF in our analysis is 2.788.¹⁰

4.3. Results

The results of the regression analysis are presented in Table 5. Because of missing values, we lose four observations, and the analysis is based on 97 responses.

¹⁰ Hair et al. (1998) suggest that as a rule of thumb, the VIF should not exceed a value of 10; others use a stricter threshold of 6 or 7 (Cohen et al., 2003).

The model explains 28% of the variance in performance (adjusted $R^2 = 0.20$). Consistent with our first hypothesis, we observe a positive and significant direct effect of contractibility on performance ($p = 0.03$). We find that using the system for operational purposes has no effect on performance, but we do observe performance effects related to incentive-oriented and exploratory uses. Using the system for incentive purposes appears to affect performance negatively ($p = 0.02$). In support of our second hypothesis, we find that this effect becomes less negative when contractibility increases ($p = 0.06$). To further examine this finding, we rerun our model using a subsample of units with high levels of contractibility. In this additional analysis (details not reported here), we find that an incentive-oriented use of the performance measurement system has no effect on performance. These results suggest that an incentive-oriented use does not improve public sector performance universally, and will harm organizational performance in low-contractibility settings.

We also find that an exploratory use of the performance measurement system tends to improve performance ($p < 0.01$). Contrary to our third hypothesis, this positive effect appears to exist independent of the level of contractibility, for the interaction between exploratory use and contractibility is not significant. Accordingly, even when contractibility is high, an exploratory way of using the system contributes to performance. Finally, organizational and unit size do not affect performance, and there are no branch effects.

4.4. Robustness checks and additional analyses

4.4.1. Alternative measures of performance measurement system use

Even though the analysis reported in Table 5 shows no signs of problematic multicollinearity, a potential concern with the analysis may still be the extent of the empirical

Table 5
OLS regression results.

Dependent variable:PERFORM	Predicted sign	Coefficient	Std. error	t-Statistic	Prob. (2-tailed)
Intercept		1.94	0.48	4.03	<0.01
INC-USE	Not predicted	-0.16	0.07	-2.31	0.02
EXPL-USE	Not predicted	0.38	0.10	3.73	<0.01
CONTRACT	H1: +	0.08	0.038	2.19	0.03
CONTRACT × INC-USE	H2: +	0.12	0.06	1.89	0.06
CONTRACT × EXPL-USE	H3: -	0.02	0.06	0.42	0.68
OP-USE	Not predicted	-0.13	0.10	-1.20	0.23
SIZEORG	Not predicted	0.03	0.10	0.35	0.73
SIZEUNIT	Not predicted	0.04	0.09	0.46	0.65
CG	Not predicted	0.11	0.11	1.03	0.31
LG	Not predicted	0.17	0.13	1.24	0.21
R-squared		0.2848			
Adjusted R-squared		0.2016			
F-statistic		3.4245			
Prob. (F-statistic)		<0.01			

PERFORM = unit performance; CONTRACT = contractibility (clear and measurable goals, knowledge of transformation processes); OP-USE = operational use of performance measures; INC-USE = incentive-oriented use of performance measures; EXPL-USE = exploratory use of performance measures; SIZEORG = size organization (=log fte organization); SIZEUNIT = size unit (=log fte of the unit); CG = central government (dummy variable); LG = local government (dummy variable).

Table 6
Regression results: uses based on residuals.

Dependent variable:PERFORM	Predicted sign	Coefficient	Std. error	t-Statistic	Prob. (2-tailed)
Intercept		2.26	0.53	4.22	<0.01
INC-USE(res)	Not predicted	-0.08	0.08	-1.08	0.28
EXPL-USE(res)	Not predicted	0.33	0.12	2.79	0.01
CONTRACT	H1: +	0.08	0.04	1.97	0.05
CONTRACT × INC-USE(res)	H2: +	0.09	0.05	1.66	0.10
CONTRACT × EXPL-USE(res)	H3: -	0.02	0.06	0.36	0.72
OP-USE	Not predicted	0.06	0.09	0.65	0.52
SIZEORG	Not predicted	0.03	0.10	0.27	0.79
SIZEUNIT	Not predicted	0.05	0.09	0.56	0.58
CG	Not predicted	0.09	0.11	0.82	0.41
LG	Not predicted	0.14	0.14	1.02	0.31
R-squared		0.2464			
Adjusted R-squared		0.1588			
F-statistic		2.8123			
Prob. (F-statistic)		<0.01			

PERFORM = unit performance; INC-USE(res) = Incentive use residual (residual from regressing INC-USE on operational use and exploratory use); EXPL-USE(res) = Exploratory use residual (residual from regressing EXPL-USE on operational use and incentive-oriented use); CONTRACT = contractibility (clear and measurable goals, knowledge of transformation processes); OP-USE = operational use of performance measures; SIZEORG = size organization (= log fte organization); SIZEUNIT = size unit (= log fte of the unit); CG = central government (dummy variable); LG = local government (dummy variable).

overlap between the constructs used to measure the various roles of the performance measurement system (OP-USE, INC-USE, and EXPL-USE).¹¹ Because of this overlap, it is conceivable that the results of the analysis are influenced by common variance, i.e. variance that is shared by the various roles of the performance measurement system. To address this concern, we redefine incentive-oriented use as the residual from a regression of the original variable INC-USE on operational use and exploratory use. Similarly, we redefine exploratory use as the residual from the regression of EXPL-USE on operational use and incentive-oriented use. These newly created variables (labelled INC-USE(res) and EXPL-USE(res), respectively) contain only variance that is unique to the individual uses, and can be said to capture only the truly distinctive components of incentive-oriented

and exploratory use (cf. Hansen and Van der Stede, 2004, for a similar approach). With these newly defined variables, we rerun the regression analysis. The results are reported in Table 6. In this additional analysis, the negative main effect of incentive-oriented use is no longer significant, but all other inferences are unaffected. This reinforces the earlier conclusion that incentive-oriented use of the performance measurement system does not generally contribute to performance, but that this effect depends on contractibility.

As a further robustness check, we rerun the analysis with another set of alternative proxies for the variables expressing performance measure use. Initially, we measured the three uses by including information on the application of five different categories of performance metrics, i.e. input, process, output, quality, and outcome measures (see Section 3.2). In the new analysis, we only look at the way in which the organizational unit uses its output and outcome measures, ignoring the information on the other categories. Arguably, these newly created

¹¹ As reported in Section 4.1, the bivariate correlations between these variables are quite high, ranging from 0.521 to 0.741.

Table 7
Regression results quantitative and qualitative performance.

	Predicted sign	Quantitative performance				Qualitative performance			
		Coefficient	Std. error	t-Statistic	Prob. (2-tailed)	Coefficient	Std. error	t-Statistic	Prob. (2-tailed)
Intercept		2.17	0.48	4.55	<0.01	1.77	0.57	3.12	<0.01
INC-USE	Not predicted	-0.18	0.08	-2.37	0.02	-0.14	0.08	-1.89	0.06
EXPL-USE	Not predicted	0.31	0.12	2.66	0.01	0.43	0.11	3.82	<0.01
CONTRACT	H1: +	0.08	0.05	1.77	0.08	0.08	0.04	2.12	0.04
CONTRACT × INC-USE	H2: +	0.15	0.08	1.98	0.05	0.09	0.06	1.47	0.15
CONTRACT × EXPL-USE	H3: -	0.01	0.07	0.09	0.93	0.04	0.07	0.55	0.58
OP-USE	Not predicted	-0.05	0.11	-0.46	0.65	-0.18	0.12	-1.53	0.13
SIZEORG	Not predicted	0.00	0.10	0.01	0.99	0.06	0.11	0.52	0.60
SIZEUNIT	Not predicted	-0.02	0.11	-0.23	0.82	0.09	0.10	0.91	0.37
CG	Not predicted	0.18	0.12	1.47	0.14	0.06	0.12	0.47	0.64
LG	Not predicted	0.09	0.15	0.60	0.55	0.22	0.15	1.44	0.15
R-squared		0.2198				0.2665			
Adjusted R-squared		0.1291				0.1812			
F-statistic		2.4227				3.1248			
Prob. (F-statistic)		<0.01				<0.01			

Quantitative performance = unit quantitative performance (comprising productivity, the attainment of production or service levels goals, and efficiency); Qualitative performance = unit quality performance (comprising quality or accuracy of work produced, number of innovations, process improvements or new ideas, reputation for work excellence, and morale of unit personnel); CONTRACT = contractibility (clear and measurable goals, knowledge of transformation processes); OP-USE = operational use of performance measures; INC-USE = incentive-oriented use of performance measures; EXPL-USE = exploratory use of performance measures; SIZEORG = size organization (= log fte organization); SIZEUNIT = size unit (= log fte of the unit); CG = central government (dummy variable); LG = local government (dummy variable).

variables are more closely aligned with NPM-thinking, which emphasizes results accountability and, consequently, propose outputs and outcomes as the main focus of control. The results of this alternative analysis (not reported here), however, do not change any of our substantive findings, and our conclusions remain the same.

4.4.2. Alternative operationalizations of contractibility

As discussed in Section 3.2, we measure contractibility by summing the scores on its formative indicators CLEARGOALS, MEASGOALS, and TRANSFORM. To examine the robustness of the results to alternative operationalizations of contractibility, we redefine CONTRACT as the product of the formative indicators, and rerun the model. The results of this additional analysis (not tabulated) do not affect our earlier inferences. We also estimate a partial least squares (PLS) model. PLS estimates the paths between CONTRACT and its formative indicators as part of the measurement model, providing a data-driven solution that requires no ex ante assumptions as to relative weights of the formative indicators. The PLS results (not tabulated) are qualitatively similar to the results obtained earlier, and do not change the substantive findings.

4.4.3. A further examination of the performance effects

In the original analysis reported in Table 5, performance (PERFORM) has been measured with an instrument that captures both quantitative (e.g. productivity and efficiency) and qualitative (e.g. quality of services) dimensions of unit performance. It may, however, be the case that performance measurement practices affect the quantitative and qualitative dimensions differently. Literature suggests that the effect of performance measurement on quantitative aspects of performance is likely to be positive, yet this positive effect may come at the expense of qualitative aspects of performance (cf. Jenkins et al., 1998, for a meta overview

of the relevant literature, and Verbeeten, 2008, for a more elaborate discussion in a specific public sector context).

To examine whether such differences exist within our dataset, we split the original performance variable in two new variables expressing quantitative performance respectively qualitative performance. Quantitative performance comprises productivity, the attainment of production or service levels goals, and efficiency. Qualitative performance includes the quality or accuracy of work produced, the number of innovations, process improvements, or new ideas, the reputation for work excellence, and the morale of unit personnel (cf. Verbeeten, 2008). Both new metrics have acceptable Cronbach's alphas (0.669 and 0.649, respectively). We rerun the regression with these new performance dimensions as the dependent variables; the results are presented in Table 7. All results remain substantially the same, albeit that in the model with qualitative performance as the dependent variable, the interaction effect of incentive-oriented use and contractibility is only one-tailed significant. This additional analysis strengthens our earlier inference: positive effects of an incentive-oriented use of performance measures can only be realized in conditions of high contractibility, even if the goal is to improve quantitative performance.

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). While the importance of

clear and measurable objectives is recognized by NPM-advocates, it is generally seen as a choice variable: if goals are ambiguous, (political) management should make them clear (cf. Dewatripont et al., 1999; Hood, 1995; Wilson, 1989). It is undoubtedly true that goal ambiguity is deliberately created in some public sector organizations for political or self-serving purposes (Kerr, 1975; Hofstede, 1981). But NPM seems to ignore that contractibility is often not a choice variable at all, and that objectives often are inherently ambiguous in public sector organizations (cf. Burgess and Ratto, 2003; Cavalluzzo and Ittner, 2004; Dixit, 1997, 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. As a result, the organization gains coherence and focus in pursuit of its mission, energizing intrinsically motivated public sector managers and employees. But it is also conceivable that we have overstated the relative costs of an exploratory use of the performance measurement system and that in general, the benefits of exploratory use outweigh the cost. Future studies could seek to address this issue.

As in any study, there are several limitations. Our data came exclusively from Dutch public sector organizations. Although the limited emphasis on incentive-focused performance contracting is characteristic of the public sector at large,¹³ it is nevertheless conceivable that some of our results are affected by the context in which they were obtained. The Dutch public sector is characterized by a general reluctance to the 'naming and shaming' of low performing organizations and managers, which may dilute formal links between performance and compensation. For instance, human resource policies in the Dutch public sector typically stipulate that employees can only qualify for in-range salary increases if their performance is at least satisfactory. In practice, however, such increases are often treated as a right that can only be denied if performance is exceptionally bad. In our discussion above, we have already suggested that this specific institutional setting—and particularly its consequences for our sample—might drive some of our results. We have no data to examine the extent to which this context differs across nations or regions of the world. An informed guess could be that similar settings can be found in for instance Sweden, Finland and some continental European countries, but that Anglo-Saxon countries are less hesitant to identify weak performers (cf. Pollitt and Bouckaert, 2011). Future studies could examine whether our findings generalize to different institutional settings.

Other caveats include the following. First, the results presented here are based on perceptions rather than 'hard' measures. Perceptions may be biased, and our proxies may misrepresent some of the key concepts in the analyses because of inappropriate measures or inadequate interpretation of the survey instrument (cf. Ittner and Larcker, 2001). For instance, we rely exclusively on self-reported performance, and we do not investigate whether—for example—politicians or the general public are happy with

¹² The mean score for incentive-oriented use of the performance measurement system is only 2.2 on a five-point scale, with a standard deviation of 0.8 (see Table 4). This implies that roughly 80% of the respondents report scores on this type of use that are equal to or lower than 3. Only 4% of the respondents score the intensity of incentive-oriented use at 4 or higher. In the semantic scale we use, a score of 3 means that the respondent uses his or her system 'sometimes' for incentive purposes.

¹³ In most OECD public services, variable performance related pay provides less than 5% of total compensation. The Dutch figure is about 3% (OECD, 2007). This figure excludes in-range salary promotions, which may be linked to performance.

the results achieved by the organizational unit. While our reliance on validated instruments and the pre-test of the questionnaire should alleviate such concerns, additional research is necessary to bolster the results from this study. Second, our model is relatively simple, and additional factors such as behavioural and cultural controls, differences in the allocation of decision rights, or mutual trust among stakeholders and managers may affect the use and effects of performance measurement systems. These additional factors have not been investigated in this study.

Despite these limitations, our study seems to have some interesting implications, both for practice and for public policy. Performance measurement systems are an important component of the management control structure. In designing such systems, public sector managers need to consider not only what to measure and how to measure it, but also how they are going to use the performance information, given the circumstances they face. Our study documents that such managerial decisions have a significant impact on subsequent performance. This is a delicate balance to strike, as our results suggest that the different purposes for using performance measures are associated. As for public policy, attempting to improve public sector performance is undoubtedly important. NPM-type performance contracting with its emphasis on incentives and individual accountability, however, is not the universal answer. In fact, our evidence shows that an incentive-oriented use à la NPM may actually have dysfunctional consequences. Therefore, politicians and policy makers need to broaden their repertoire beyond conventional NPM thinking in order to improve the effectiveness of the public sector.

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.

Appendix.

Survey (abbreviated)

A.1. Performance (PERFORM)

How do you rate the performance of your unit relative to other, comparable units (in the public, but perhaps also in the private sector) on each of the following dimensions? (1 = far below average, 5 = far above average)

- a. The amount of work and/or the number of products produced in your unit.
- b. The accuracy of work produced in your unit and/or the quality of the goods delivered.
- c. The number of innovations, process improvements, or new ideas implemented by your unit.
- d. The reputation for work excellence of your unit.
- e. The attainment of production or service level goals of your unit.
- f. The efficiency of operations within your unit.
- g. The morale of unit personnel.

A.2. Contractibility (CONTRACT)

Clarity of mission and goals (CLEARGOALS)

Based on your experience, please indicate the extent to which you agree to the following statements. (1 = strongly disagree, 5 = strongly agree)

- a. My unit's mission is unequivocal.
- b. My unit's mission is written on paper and is communicated internally and externally.
- c. My unit's goals are unambiguously related to the mission.
- d. The set of goals of my unit is internally inconsistent (reverse coded).
- e. My unit's goals are specific and detailed.
- f. My unit's goals change regularly as a result of political developments (reverse coded).

Note: items d and f were dropped from the analysis because of insufficient factor loading.

Measurability of outputs (MEASGOALS)

Based on your experience, please indicate the extent to which you agree to the following statements. (1 = strongly disagree, 5 = strongly agree)

- a. The goals of my unit are expressed in a wholly quantitative way (e.g. budget, productivity, numbers).
- b. Intended performance for my unit is expressed in no more than five performance metrics.
- c. The set of performance metrics provides a complete picture of the results to be achieved.
- d. The performance measures of the unit are unambiguously related to the goals of the organization
- e. Goal achievement depends heavily on external factors (reverse coded).
- f. The causal relation between resource allocation and goal achievement is absolutely clear.
- g. The effect of our efforts become visible within a year.

Note: items a and b were dropped from the analysis because of insufficient factor loading.

Knowledge of transformation processes (TRANSFORM)

Based on your experience, please indicate the extent to which you agree to the following statements. (1 = strongly disagree, 5 = strongly agree)

- a. In performing our tasks, there is a logical way to proceed.
- b. The unit's primary processes can only be performed in one specific and documented way.
- c. Within the unit, it is entirely clear how to perform our tasks.
- d. In performing their tasks, unit employees rely on standard procedures and rules.

A.3. Performance measurement system use (OP-USE, INC-USE, EXPL-USE)

For which purposes do you use (input/process/output/quality/outcome) measures?

- a. Operational planning.
- b. Budget allocation.
- c. Process monitoring.
- d. Revision of business processes.
- e. Career decisions regarding individual employees.
- f. Bonus determination.
- g. Communicating goals and priorities to unit employees.
- h. Evaluating the appropriateness of goals and/or policy assumptions.
- i. Revision of the unit's policy.
- j. Reporting to senior management.
- k. Reporting to external stakeholders.

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.

References

- Abernethy, M.A., Brownell, P., 1999. *The role of budgets in organizations facing strategic change: an exploratory study*. *Accounting, Organizations and Society* 24, 189–204.
- Ahrens, T., Chapman, C.S., 2004. *Accounting for flexibility and efficiency: a field study of management control systems in a restaurant chain*. *Contemporary Accounting Research* 21 (2), 271–301.
- Aiken, L.S., West, S.G., 1991. *Multiple Regression: Testing and Interpreting Interactions*. Sage Publications, Newbury Park.
- Argyris, C., 1977. *Organizational learning and management information systems*. *Accounting, Organizations and Society* 2, 113–123.
- Argyris, C., 1990. *The dilemma of implementing controls: the case of managerial accounting*. *Accounting, Organizations and Society* 15, 503–511.
- Baker, G., 2002. *Distortion and risk in optimal incentive contracts*. *Journal of Human Resources* 37, 728–751.
- Bevan, G., Hood, C., 2006. *What's measured is what matters: targets and gaming in the English public health care system*. *Public Administration* 84, 517–538.
- Bisbe, J., Batista-Foguet, J., Chenhall, R., 2007. *Defining management accounting constructs: a methodological note on the risks of*

- conceptual misspecification*. *Accounting, Organizations and Society* 32, 789–820.
- Brignall, S., Modell, S., 2000. *An institutional perspective on performance measurement and management in the 'new public sector'*. *Management Accounting Research* 11, 281–306.
- Broadbent, J., Laughlin, R., 1998. *Resisting the 'new public management': absorption and absorbing groups in schools and GP practices in the UK*. *Accounting, Auditing and Accountability Journal* 11, 403–435.
- Burchell, S., Clubb, C., Hopwood, A., Hughes, J., 1980. *The roles of accounting in organizations and society*. *Accounting, Organizations and Society* 5, 5–27.
- Burgess, S., Ratto, M., 2003. *The role of incentives in the public sector: issues and evidence*. *Oxford Review of Economic Policy* 19, 285–299.
- Campbell, D., Epstein, M., Martinez-Jerez, A., 2008. *Implicit Control Mechanisms and Organizational Learning*. Working paper.
- Cavalluzzo, K.S., Ittner, C.D., 2004. *Implementing performance measurement innovations: evidence from government*. *Accounting, Organizations and Society* 29, 243–267.
- Chun, Y.H., Rainey, G.H., 2005. *Goal ambiguity and organizational performance in U.S. federal agencies*. *Journal of Public Administration Research and Theory* 15, 529–557.
- Churchill, G.A., 1979. *A paradigm for developing better measures of marketing constructs*. *Journal of Marketing Research* 16, 64–73.
- Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, third ed. Lawrence Erlbaum Associates, Mahwah.
- De Bruijn, H., 2002. *Performance measurement in the public sector: strategies to cope with the risks of performance measurement*. *International Journal of Public Sector Management* 15, 578–594.
- De Bruijn, H., Van Helden, G.J., 2006. *A plea for dialogue driven performance-based management systems: evidence from the Dutch public sector*. *Financial Accountability & Management* 22 (4), 405–423.
- Demski, J.S., Feltham, G.A., 1976. *Cost Determination: A Conceptual Approach*. Iowa State University Press, Ames.
- Dewatripont, M., Jewitt, I., Tirole, J., 1999. *The economics of career concerns, part II: application to missions and accountability of government agencies*. *Review of Economic Studies* 66, 199–217.
- Diamantopoulos, A., Sigauw, J.A., 2006. *Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration*. *British Journal of Management* 17, 263–282.
- Dixit, A., 1997. *Power of incentives in private versus public organizations*. *American Economic Review* 87, 378–382.
- Dixit, A., 2002. *Incentives and organizations in the public sector: an interpretive review*. *The Journal of Human Resources* 37, 696–718.
- Dunk, A.S., Lysons, A.F., 1997. *An analysis of departmental effectiveness, participative budgetary control processes and environmental dimensionality within the competing values framework: a public sector study*. *Financial Accountability & Management* 13, 1–15.
- Echambadi, R., Hess, J.D., 2007. *Mean-centering does not alleviate collinearity problems in moderated multiple regression models*. *Marketing Science* 26, 438–445.
- Echambadi, R., Campbell, B., Agerwal, R., 2006. *Encouraging best practice in quantitative management research: an incomplete list of opportunities*. *Journal of Management Studies* 43, 1801–1820.
- Feltham, G.A., Xie, J., 1994. *Performance measure congruity and diversity in multi-task principal/agent relations*. *The Accounting Review* 69, 429–453.
- Franco-Santos, M., Kennerley, M., Micheli, P., Martinez, V., Mason, S., Marr, B., Gray, D., Neely, A., 2007. *Towards a definition of a business performance measurement system*. *International Journal of Operations & Production Management* 27, 784–801.
- Gatignon, H., Vosgerau, J., 2005. *Moderating effects: the myth of mean centering*. INSEAD.
- Gibbons, R., 1998. *Incentives in organizations*. *Journal of Economic Perspectives* 12 (4), 115–132.
- Goddard, A., 2010. *Contemporary public sector accounting research: an international comparison of journal papers*. *British Accounting Review* 42, 75–87.
- Goold, M., Quinn, J.J., 1990. *The paradox of strategic controls*. *Strategic Management Journal* 11, 43–57.
- Grafton, J., Lillis, A.M., Widener, S.K., 2010. *The role of performance measurement and evaluation in building organizational capabilities and performance*. *Accounting, Organizations and Society* 35, 689–706.
- Groot, T., Budding, T., 2008. *New public management's current issues and future prospects*. *Financial Accountability & Management* 24, 1–13.
- Groot, T.L.C.M., Van Helden, G.J., 2012. *Financieel Management Van Non-profit Organisaties*. Wolters-Noordhoff, Groningen.

- Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. *Multivariate Data Analysis*, fifth ed. Prentice Hall, Upper Saddle River.
- Hall, M., 2008. The effect of comprehensive performance measurement systems on role clarity, psychological empowerment and managerial performance. *Accounting, Organizations and Society* 33, 141–163.
- Hansen, S.C., Van der Stede, W.A., 2004. Multiple facets of budgeting: an exploratory analysis. *Management Accounting Research* 15, 415–439.
- Hartmann, F.G.H., Moers, F., 1999. Testing contingency hypotheses in budgetary research: an evaluation of the use of moderated regression analysis. *Accounting, Organizations and Society* 24, 291–315.
- Heinrich, C.J., 2002. Outcomes-based performance management in the public sector: implications for government accountability and effectiveness. *Public Administration Review* 62, 712–725.
- Henri, J.F., 2006. Organizational culture and performance measurement systems. *Accounting, Organizations and Society* 31, 77–103.
- Hofstede, G., 1981. Management control of public and not-for-profit activities. *Accounting, Organizations and Society* 6, 193–211.
- Holmstrom, B., 1982. Moral hazard in teams. *Bell Journal of Economics* 13, 324–340.
- Hood, C., 1995. The 'new public management' in the 1980: variations on a theme. *Accounting, Organizations and Society* 20, 93–109.
- Hood, C., 2007. Public service management by numbers: why does it vary? Where has it come from? What are the gaps and the puzzles? *Public Money & Management* 27 (2), 95–102.
- Hyndman, N., Eden, R., 2000. A study of the coordination of mission, objectives and targets in U.K. executive agencies. *Management Accounting Research* 11, 175–191.
- Ittner, C.D., Larcker, D.F., 2001. Assessing empirical research in management accounting: a value-based perspective. *Journal of Accounting and Economics* 32, 349–410.
- Jaccard, J., Turrisi, R., Wan, C.K., 1990. *Interaction Effects in Multiple Regression*. Sage Publications, Newbury Park.
- Jarvis, C.B., Mackenzie, S.B., Podsakoff, P.M., 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research* 30, 199–218.
- Jenkins Jr., G.D., Mitra, A., Gupta, N., Shaw, J.D., 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology* 83, 777–787.
- Jung, C.S., Rainey, H.G., 2008. Developing the concept of program goal ambiguity and explaining federal program performance. *Academy of Management Proceedings*.
- Kelman, S., Friedman, J.N., 2009. Performance improvement and performance dysfunction: an empirical examination of distortionary impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory* 19, 917–946.
- Kerr, S., 1975. On the folly of rewarding A, while hoping for B. *Academy of Management Journal* 18, 769–783.
- Lee, F., Edmonson, A.C., Thomke, S., Worline, M., 2004. The mixed effects of inconsistency on experimentation in organizations. *Organization Science* 15, 310–326.
- Locke, E.A., Latham, G.P., 2002. Building a practically useful theory of goal setting and task motivation: a 35 year odyssey. *American Psychologist* 57, 705–717.
- Marsden, D., Belfield, R., 2006. Pay for performance where output is hard to measure: the case of performance pay for school teachers. In: Kaufman, B.E., Lewin, D. (Eds.), *Advances in Industrial and Labor Relations*, vol. 15. JAI Press, London, pp. 1–37.
- Mundy, J., 2010. Creating dynamic tensions through a balanced use of management control systems. *Accounting, Organizations and Society* 35 (5), 499–523.
- Naranjo-Gil, D., Hartmann, F., 2007. Management accounting systems, top management team heterogeneity, and strategic change. *Accounting, Organizations and Society* 32, 735–756.
- Newberry, S., Pallot, J., 2004. Freedom or coercion? NPM incentives in New Zealand central government departments. *Management Accounting Research* 15, 247–266.
- OECD, 2007. Performance-based arrangements for senior civil servants: OECD and other country experiences, OECD working papers on public governance 2007/5.
- Otley, D.T., Berry, A.J., 1980. Control, organization and accounting. *Accounting, Organizations and Society* 5, 231–244.
- Ouchi, W.G., 1977. The relationship between organizational structure and organizational control. *Administrative Science Quarterly* 22, 95–113.
- Pollitt, C., 2002. Clarifying convergence: striking similarities and durable differences in public management reform. *Public Management Review* 4, 471–492.
- Pollitt, C., 2006. Performance management in practice: a comparative study of executive agencies. *Journal of Public Administration Research and Theory* 16 (1), 25–44.
- Pollitt, C., Bouckaert, G., 2011. *Public Management Reform, a Comparative Analysis: New Public Management, Governance, and the Neo-Weberian State*, third ed. Oxford University Press, Oxford/New York.
- Prendergast, C., 1999. The provision of incentives in firms. *Journal of Economic Literature* 37, 7–63.
- Propper, C., Wilson, D., 2003. The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy* 19, 250–267.
- Roberts, J., 2004. *The Modern Firm: Organizational Design for Performance and Growth*. Oxford University Press, Oxford.
- Simons, R., 1990. The role of management control systems in creating competitive advantage: new perspectives. *Accounting, Organizations and Society* 15, 127–143.
- Simons, R., 2000. *Performance Measurement and Control Systems for Implementing Strategy*. Prentice Hall, Upper Saddle River, NJ.
- Speklé, R.F., 2001. Explaining management control structure variety: a transaction cost economics perspective. *Accounting, Organizations and Society* 26, 419–441.
- Sprinkle, G.B., 2003. Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society* 28, 287–318.
- Sprinkle, G.B., Williamson, M.G., Upton, D.R., 2008. The effort and risk-taking effects of budget-based contracts. *Accounting, Organizations and Society* 33, 436–452.
- Taylor, J., 2011. Strengthening the link between performance measurement and decision making. *Public Administration* 89, 860–878.
- Ter Bogt, H.J., 2008. Management accounting change and new public management in local government: a reassessment of ambitions and results – An institutionalist approach to accounting change in the Dutch public sector. *Financial Accountability & Management* 24, 209–241.
- Ter Bogt, H., Budding, T., Groot, T., Van Helden, J., 2010. Current NPM research: digging deeper and looking further. *Financial Accountability & Management* 26, 241–245.
- Ter Bogt, H.J., Van Helden, G.J., 2005. A reflection on accounting reforms in Dutch government. In: Guthrie, J., Humphrey, C., Jones, L.R., Olson, O. (Eds.), *International Public Financial Management Reform: Progress, Contradictions, and Challenges*. Information Age Publishing Inc, Greenwich, CT, USA.
- Tessier, S., Otley, D., 2012. A conceptual development of Simons' levers of control framework. *Management Accounting Research* 23, 171–185.
- Tirole, J., 1994. The internal organization of government. *Oxford Economic Papers* 46, 1–29.
- Van de Ven, A.H., Ferry, D.L., 1980. *Measuring and Assessing Organizations*. Wiley, New York.
- Van Helden, G.J., 1998. A review of the policy and management instruments project for municipalities in the Netherlands. *Financial Accountability & Management* 14, 85–104.
- Van Helden, G.J., 2005. Researching public sector transformation: the role of management accounting. *Financial Accountability & Management* 21, 99–133.
- Van Veen-Dirks, P., 2010. Different uses of performance measures: the evaluation versus reward of production managers. *Accounting, Organizations and Society* 35, 141–164.
- Verbeeten, F.H.M., 2008. Performance management practices in public sector organizations: impact on performance. *Accounting, Auditing and Accountability Journal* 21, 427–454.
- Walker, R.W., Damanpour, F., Devece, C.A., 2011. Management innovation and organizational performance: the mediating effect of performance management. *Journal of Public Administration Research and Theory* 21, 367L 86.
- Whithey, M., Daft, R.L., Cooper, W.H., 1983. Measures of Perrow's work unit technology: an empirical assessment and a new scale. *Academy of Management Journal* 26, 45–63.
- Williams, J.J., Macintosh, N.B., Moore, J.C., 1990. Budget-related behavior in public sector organizations: some empirical evidence. *Accounting, Organizations and Society* 15, 221–246.
- Wilson, J., 1989. *Bureaucracy*. Basic Books, New York.