Nondyadic control systems and effort direction effectiveness: Evidence from the public sector

Speklé, R.F.; Verbeeten, F.H.M.; Widener, S.K.

DOI
10.1016/j.mar.2021.100769

Publication date
2022

Document Version
Final published version

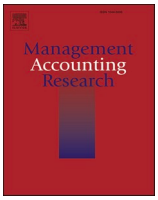
Published in
Management Accounting Research

Management Accounting Research 54 (2022) 100769

Contents lists available at ScienceDirect

Management Accounting Research

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



Nondyadic control systems and effort direction effectiveness: Evidence from the public sector

Roland F. Speklé [a,*], Frank H.M. Verbeeten [b], Sally K. Widener [c]

[a] Nyenrode Business University, Netherlands
[b] University of Amsterdam, Netherlands
[c] Clemson University, United States

ARTICLE INFO

Keywords:
Performance measurement
Contractibility
Systems of control
Complementarities
Public Sector

ABSTRACT

One of the main themes in the current management control literature is the study of systems of control practices. While several studies have examined pairwise ('dyadic') complementarities in control practices to investigate whether a control system exists, control systems may also be comprised of more than two control practices ('nondyadic systems'). We examine whether organizations use a complementary system of three performance measurement uses to mitigate the control problem of effort direction. We illustrate how a nondyadic control system can be modeled and empirically examined using data from 162 organizational units in the public sector. Our findings are consistent with our hypothesis that operational, incentive-oriented, and exploratory uses of performance information are complementary and combine in a single system in a low contractibility setting. We also show that the intensity of use of this nondyadic control system is significantly correlated with effort direction effectiveness in conditions of low contractibility. In contrast, and consistent with our theory, we find that when contractibility is high, the combination of the three uses of performance measures is not correlated with effort direction. Jointly, our results suggest that multiple performance measurement uses combine as a nondyadic system to guide effort direction in a low contractibility setting, yet not in a high contractibility setting.

1. Introduction

One main purpose of management control is to provide direction to organizational members so they can channel their efforts in desired ways (Merchant and Van der Stede, 2017). Such effort direction is relatively easy to achieve in the presence of a well-developed predictive model of the outcomes, activities, and processes involved in goal achievement (Emmanuel et al., 1990; Otley and Berry, 1980). In such conditions, desired contributions to the organization can be specified in advance in a reasonably complete contract between principal and agent; we refer to this situation as a high contractibility setting (e.g., Moers, 2006; Ouchi, 1977; Speklé and Verbeeten, 2014). However, when contractibility is low, providing direction is much more difficult and a more intricate response from the organization is required to solve the effort direction problem.

In this paper, we propose that such a response involves a system comprised of the joint use of performance measures for operational, incentive-oriented, and exploratory purposes. We refer to a system that involves more than two interrelated control practices as a 'nondyadic system'.[1] An operational use of performance measures provides organizational members with information on routine decision making and on-going operational processes. An incentive-oriented use provides a basis for accountability, rewards, and evaluation processes. An exploratory use facilitates learning and exploring in the organization. We argue that, in conditions of low contractibility, all three uses are necessary and must be used jointly because they complement and require each other to contribute to the solution of the control problem of effort direction. We furthermore argue that, when contractibility is high, the three performance measurement practices do not combine in a single system that addresses the aforementioned effort direction challenges.

We examine our hypothesis in three steps. First, following recent empirical studies (e.g., Abernethy et al., 2015; Grabner, 2014), we regress each performance measurement use on a series of contingency factors that may jointly affect these uses. This removes any shared

* Corresponding author.
E-mail address: r.spekle@nyenrode.nl (R.F. Speklé).

[1] Deferring a more in-depth discussion to Section 2.2, we define a nondyadic system as a system comprising three or more control practices in which all practices are interrelated, either directly or indirectly.

https://doi.org/10.1016/j.mar.2021.100769
Received 8 April 2020; Received in revised form 10 July 2021; Accepted 4 September 2021
Available online 11 September 2021

variance in the performance measurement uses due to joint determinants. Second, we then use the regression residuals in a common factor analysis to estimate the remaining shared variance among the multiple performance measurement uses, conditional on contractibility. The emergence of a single factor provides evidence that the performance measurement uses work in combination but does not, in and by itself, provide evidence on whether those systemic effects address the specific control problem of effort direction. Rather, the use of factor analysis may be considered as an empirical test of a necessary yet not sufficient condition; if a system exists, the practices must share part of their variance which we will pick up in a common factor analysis, yet finding a factor is in itself not sufficient to indicate the existence of a system of the performance measurement uses that addresses effort direction. Therefore, we subsequently estimate conditional correlations in the third step of our analysis to investigate whether a higher reliance on the combination of performance measure uses (i.e., a higher score on the factor representing the combination of control practices) is (is not) on average associated with higher performance on the specified control problem, i.e., effort direction, in conditions of low (high) contractibility. This performance specification provides evidence that the combination of performance measurement uses is indeed related to the specific problem theorized (in this study: effort direction), rather than to some alternative control problem.

We test our hypothesis in a sample of 162 organizational units in the Dutch public sector. Effort direction problems are quite real in the public sector, in which goals are often ambiguous, outcomes are relatively difficult to measure, and knowledge of the transformation process tends to be imperfect (Burgess and Ratto, 2003; Hofstede, 1981; Tirole, 1994). Previous work has found that both contractibility levels and the structural response to the associated control problem of effort direction are likely to vary considerably within the Dutch public sector (Speklé and Verbeeten, 2014); thus, this setting is relevant and informative for our purposes.

Our empirical findings are consistent with our hypothesis. We find that the combination of the three performance measurement uses is correlated with effort direction effectiveness in a low contractibility setting. In the high contractibility context, we also find some interrelatedness among the various performance measurement uses; however, the combined use of performance measures is not correlated with effort direction effectiveness in this latter setting. Thus, our factor and conditional correlation analyses suggest that the three performance measurement uses comprise a nondyadic system of controls in which the three practices should be applied together to solve the control problem of effort direction when contractibility is low.

Our study makes three main contributions to the literature. First, we contribute to the literature on control practices as a system (Grabner and Moers, 2013; Malmi and Brown, 2008; Hofmann and Van Lent, 2017; Masschelein and Moers, 2020). We provide additional insights into the relationship between contractibility, associated control problems, control choices, and their joint effect on the functioning of organizations. Contractibility and its underlying dimensions of goal clarity, measurability of outputs, and knowledge of the transformation process have been studied extensively as antecedents of a variety of different individual control practices, albeit sometimes under different labels and in different but related manifestations.[2] Most studies in this area have focused on motivational problems (see Holmström, 2017; Otley, 2016); we extend this literature to effort direction challenges, which are not trivial in many organizations (Merchant and Van der Stede, 2017). In addition, previous studies have tended to ignore the possibility of systematic links between the individual practices to solve control problems and may consequently suffer from model misspecification (Chenhall, 2003). Our study documents that, in fact, such links amongst performance measure practices exist and, furthermore, that the performance measure practices belong to a single system that is associated with effort direction effectiveness in conditions of low contractibility (yet not in high contractibility).

Second, we contribute to the public sector literature. Public sector organizations face continuing pressure to reform their management style and practices to resemble private sector organizations more closely (Hood, 1995). This pressure, however, disregards that contractibility is typically lower in the public sector as compared to the private sector. This makes contractibility and the associated problem of effort direction especially salient for public sector organizations, driving the need for richer insights into the use of performance measure practices in low contractibility settings. Previous studies in the public sector have taken a contingency approach to show that contractibility moderates the association between incentive-oriented use of performance measurement and organizational performance (Speklé and Verbeeten, 2014). We add to this literature by showing that multiple performance measurement practices form a nondyadic system to mitigate the effort direction problem in a context of low contractibility. That is, when contractibility is low, effort direction effectiveness increases only when the combination of all three uses is intensified. Our focus, thus, is on a nondyadic system of performance measurement uses as opposed to the isolated uses of performance measures that have been the center of prior studies in this line of work.

Finally, while previous literature has mostly focused on interdependencies between two control practices (i.e., dyadic control systems), we discuss and illustrate a three-step empirical method of testing a management control system comprised of more than two control practices (i.e., a nondyadic control system). To capture the complementarities of multiple control practices, recent papers have suggested the use of common factor analysis (Nielsen et al., 2018; Speklé et al., 2017; Tanriverdi and Venkatraman, 2005). We draw on the intuition of these studies but extend this approach by controlling for a theoretically motivated set of contingency factors that may jointly affect the use of the practices and for other outcome effects that may conflate performance on the control problem the performance measurement uses are solving. Although we focus on three performance measurement practices in our empirical analysis, our method can be extended to include more than three control practices, assuming that the researcher can provide a theory that connects these practices to solve the control problem at hand (Grabner and Moers, 2013; Hofmann and Van Lent, 2017; Masschelein and Moers, 2020). One strength of the factor approach is that it accommodates the analysis of different direct and indirect patterns in which the practices can be related and does not require full specification of exactly which practices are complements.[3] Another strength is that the factor approach can pick up on higher level interactions that may be difficult to theorize and/or interpret. By illustrating this approach, we hope to spur future research that extends the complexity of the control system considered in the analysis beyond singular pairwise complementarities.

The remainder of this paper is structured as follows. In Section 2, we theorize the relationships among contractibility, effort direction, and the systemic use of performance measurement practices. In Section 3, we provide details of our sample, survey development, and constructs. Section 4 discusses our empirical approach and presents the results. Finally, in Section 5, we conclude and discuss limitations and implications.

[2] Note in this context that there is considerable conceptual overlap between contractibility and the broad umbrella-like notion of uncertainty, which in its various sub-forms may well be the most intensively studied phenomenon in the contingency literature (Chenhall, 2003).

[3] See Section 2.2 for a further elaboration on these issues.

2. Background literature and theoretical development

2.1. Contractibility and effort direction

Contractibility refers to the degree to which desired contributions to the organization can be specified in advance in a reasonably complete contract between principal and agent. Complete contracting assumes the availability of a well-developed predictive model of the processes being controlled (Emmanuel et al., 1990), which in turn implies that (1) goals and objectives can be specified unambiguously in advance; (2) the organization is able to selected undistorted performance measures, i.e., measures that provide incentives that are aligned with the organization's ultimate objectives; and (3) 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 actions (Baker, 2002; Gibbons, 1998; Hofstede, 1981; Otley and Berry, 1980; Speklé and Verbeeten, 2014). In a high contractibility setting, these cumulative requirements hold. In contrast, low contractibility refers to a situation where goals are ambiguous, performance measures are incomplete or distorted, and employees do not fully understand how their efforts will translate into results.

Merchant and Van der Stede (2017) suggest that control problems can be classified into three main categories: lack of direction (employees do not know what the organization wants from them), motivational problems (individual and organizational objectives do not naturally coincide and employees may display self-interested behavior), and personal limitations (such as a lack of aptitude, training, experience, stamina or knowledge for the tasks at hand). When lack of direction occurs, employees perform inadequately because they do not know what the organization wants from them; hence one function of management controls is to provide guidance to employees so they can direct their contributions towards the fulfillment of the organization's 'true objectives' (Merchant and Van der Stede, 2017). In a low contractibility setting, steering effort is difficult: goals are ambiguous (which makes it difficult to communicate them to employees), performance standards are difficult to quantify (as performance measures are incomplete or distorted), and alignment of operational activities with the 'true objectives' of the organization is complex (as outcomes of activities are difficult to predict). Low contractibility settings, thus, offer organizational actors little insight for effort direction choices. Therefore, effort direction is among the key control problems that need to be solved in that setting.

2.2. Nondyadic management control systems

Management control practices form a system if the individual practices are interdependent, i.e., when the value of using a particular practice depends on the simultaneous use of other practices in the system (Grabner and Moers, 2013; Hofmann and Van Lent, 2017; Masschelein and Moers, 2020). Studies in management control have generally focused on systems comprising two practices, but systems can consist of a larger number of complementary control choices (Bedford, 2020; Speklé and Widener, 2020). We use the term 'nondyadic system' to refer to combinations of three or more control practices in which all practices are interrelated, either directly or indirectly. For nondyadic systems, the specific patterns of the interrelationship between the practices can take different forms. That is, a nondyadic system that consists of three control practices can take two forms, i.e., one in which each practice is affected by both other practices in the system, and one in which two practices that hold no direct relationship to each other are nevertheless indirectly interdependent because they are both affected by the third practice. The forms of a nondyadic system can increase significantly, depending upon the number of practices considered in the empirics.

In this study, we examine the operational, incentive-oriented, and exploratory uses of performance measures. Accordingly, the form of the nondyadic system could consist of each performance measure use depending on both of the other two uses, i.e., operational use depending on the use of both incentive-oriented and exploratory uses, and so forth. However, the structure could also include indirect relationships. In this case, it is possible, for example, that operational and incentive-oriented uses depend on each other while incentive-oriented and exploratory uses also depend on each other. The three uses would then combine to form a nondyadic system even if there is no direct dependence relationship between exploratory and operational use because there is an indirect relationship between these latter uses. In this study, while our theoretical development that follows suggests a three-way complementarity, we hypothesize only that the three performance measurement uses form a system that mitigates the effort direction problem. We do not hypothesize the specifics of the internal structure of the system or provide evidence thereon. The reason for this is twofold. One, the empirical strategy used in this study does not allow us to reach definitive conclusions about the specific pattern of complementarities. Second, the form does not matter from a practical point of view because if the three performance measurement uses form a system, they should be applied together regardless of whether the system contains only direct relationships or a combination of direct and indirect relationships.[4]

2.3. Performance measure uses

Following previous literature (e.g., Speklé and Verbeeten, 2014), we distinguish three main uses of performance metrics: an operational use, an incentive-oriented use, and an exploratory use. These three different uses are not mutually exclusive; they co-exist in organizations. For example, Verbeeten and Speklé (2015) report strong correlations between various uses of performance information and suggest that these correlations may be driven by differences in management styles (i.e., a numbers-based versus an intuitive style) or spill-over effects between the uses. We develop a third potential explanation, i.e., that low contractibility settings require a nondyadic system of all three performance measurement uses wherein the contribution of each individual practice to solving the problem of effort direction depends on both other practices. This is because in a low contractibility setting, each type of performance measurement use has limitations; thus, no one performance measurement use can solve the effort direction problem. Furthermore, each individual use is also associated with potential dysfunctional consequences, and these may offset the benefits of any one performance measurement use towards solving the effort direction problem. For this reason, our theory holds no implications for any main effects of the uses, and we do not hypothesize any. However, when one performance measurement use is employed in the presence of the other two uses, we contend that the limitations of each are overcome and the dysfunctional effects are offset, thus, together the joint use of all three performance measurement uses will help solve the effort direction problem. Accordingly, we theorize that all three are needed jointly. Alternatively, in situations of high contractibility, we expect that a system of performance measures uses is not necessary to solve the effort direction challenges; as contractibility is high, dysfunctional effects of each individual use is likely to be limited. Before we can develop this hypothesis formally, though, we first discuss the benefits, limitations, and dysfunctional consequences of each performance measurement use when contractibility is low.

Organizations rely on performance measures for operational planning and resource allocation. They do so even in a low contractibility setting (Davila et al., 2009), as managers must provide at least some short-term guidance to their subordinates and ensure a minimum level of process coordination, for instance, in the form of budgets, production quantities, or available labor hours. However, because this information merely specifies the results associated with short-term operational

[4] We thank an anonymous reviewer for pointing this out.

activities and budget constraints, it only provides a partial picture of performance expectations in situations of low contractibility. Despite the basic operational information employees have received, they will still experience considerable uncertainty regarding their responsibilities and the standards used to assess their contribution to the organization. To cope with that uncertainty (Burney and Widener, 2007; Marginson et al., 2014; Marginson and Ogden, 2005), employees will search for anchors and clues that can help focus their attention and efforts (Kahn et al., 1964; Rizzo et al., 1970); as a result, the use of performance measures for operational purposes may become their 'main beacon' (Marginson and Ogden, 2005). If this happens, employees lose sight of the 'true objectives' of their organization, so the reliance on performance measures for operational purposes comes at a cost. To use a university setting to illustrate our point: while teaching allocated credit hours and meeting publication targets are important for budgetary and coordination purposes, this does not necessarily link to the true objectives of the university. These are usually much grander, and may, for instance, be formulated as 'the pursuit of world-class research that helps us understand society and its needs, and the creation of transformative learning experiences that inspire and equip people to ease these needs'.

The use of performance measures for incentive purposes is also useful for organizations that face low contractibility, as this use instills an elementary but fundamental awareness that certain targets, such as efficiency and outputs, matter, thus, reminding employees of the need to achieve results and encouraging them to be selective in their effort allocation decisions. In addition, such a basic level of evaluative pressure also serves as a catalyst for purposeful information search by employees as input for both strategic learning and operational improvements (Bedford, 2015; Chenhall and Morris, 1995). Relative to a high contractibility setting, the reliance on incentive-oriented use cannot be strong in absolute terms when contractibility is low, as that would lead to distortion of effort (Baker, 2002), neglect of unmeasured tasks (Holmstrom and Milgrom, 1991), and other dysfunctional behaviors (Feltham and Xie, 1994; Gibbons, 1998; Otley and Berry, 1980; Verbeeten and Speklé, 2015). For example, the use of student feedback for incentive purposes in universities may motivate lecturers to spend excessive time on teaching, which may damage the previously mentioned purpose of achieving research excellence. In addition, when employees' rewards become dependent on meeting targets, they will decrease their willingness to engage in experimentation, leading to premature closure of strategic learning (Argyris, 1977, 1990; Lee et al., 2004; Sprinkle et al., 2008). This is problematic if contractibility is low, as an openness to new ideas is generally essential to acquire a better understanding of decision alternatives and their potential contribution to organizational goal achievement (Burchell et al., 1980; Speklé and Verbeeten, 2014).

In a low contractibility setting, the need for experimentation and learning is high, and an exploratory use of performance is especially valuable. This way of using performance information challenges the assumptions underlying organizational decision-making, encourages discussion and debate, and stimulates double-loop learning and experimentation with new approaches and initiatives. The explorative use of performance measures provides members of the organization the opportunity to discuss alternative courses of action, as well as an opportunity to explore how such courses of action will help to achieve the overall goals. This process contributes to a shared frame of reference as to which goals should be achieved, what constitutes satisfactory performance, and what input is required to achieve the goals (Burchell et al., 1980; Gibbons and Kaplan, 2015). However, experimentation and debate are also costly, and may degenerate into a resource-consuming 'art for art's sake' exercise that takes too long to converge in sensible policy initiatives and action plans, especially when contractility is low. To continue our university example: we all know the lengthy and acrimonious discussions about the operationalization of research excellence and how best to allocate funds to schools and departments to achieve that objective. The contribution of these discussions to ultimate goal achievement is tenuous at best, and certainly consumes a lot of time from faculty and administrators that could have been put to alternative uses.

2.4. Contractibility, nondyadic systems of performance measures, and effort direction

It is clear from the discussion above that while operational, incentive, and exploratory performance measure uses each have benefits, each use also has a cost, especially in situations of low contractibility. However, regardless of the level of contractibility, organizations must provide guidance to employees regarding where to direct their efforts, and we argue that it is through the combination of all three uses that the potential dysfunctional effects associated with each individual type of use can be mitigated, while allowing the benefits of each to be realized.

Operational use provides some level of clarity on short term planning and resource allocations and needs, yet the benefit can only be realized when its downside of spurring myopic focus on certain short-term operational uses such as allocating capacity, task assignment, or budgetary allocations is mitigated. This downside can at least partially be offset when performance measurement information is used in an exploratory way that links operational performance information with the 'true objectives' of the organization while simultaneously using it an incentive-oriented way. The exploratory use forces employees to be skeptical of the stated targets, preventing employees from taking the operational measures too seriously and instead reminding them that they need to engage in learning and be open to new ideas while the incentive-oriented use provides needed focus and salience to specific objectives and goals. This requires careful calibration, because an incentive-oriented use also narrows employees' focus to achieving the stated performance goals, which are incomplete and noisy when contractibility is low. Thus, complementing the incentive use with operational use ensures that effort is also being directed towards critical short-term needs, while simultaneously the exploratory use continuously calls into question both the stated performance goals and the ideas as to how to achieve them, such that employees' efforts are not narrowly focused on only achieving incentivized targets. However, when contractibility is low, the exploratory use may degenerate into a resource-consuming and unfocused search for new ideas and solutions. Complementing the exploratory use with operational use helps rein in that experimentation and ensures that employees remain mindful of short-term objectives, while an incentive-oriented use makes clear to employees that they still must achieve some targets. In sum, when the three performance measurement complement each other, they create a control system that serves to allow the benefits of each of the performance measurement uses to be realized because their limitations are offset and their dysfunctional consequences mitigated.

To summarize our argument: in low contractibility settings, effort direction is a key control problem. In such settings, employees easily 'go off the rails,' waste time and resources, and do not make progress towards meaningful objectives. To address this problem, organizations will rely simultaneously on all three uses of performance information (operational, exploratory, and incentive-oriented use), combining them in a nondyadic system. This nondyadic system is necessary in order to prevent a more extensive use in one area (operational, exploratory or incentive-oriented use) from resulting in suboptimal effort direction, for example, focusing on short-term targets while neglecting the overall objectives of the organization. Accordingly, the benefits of every individual use are dependent on the other two uses, and a higher (lower) use for one purpose will, on average, translate into a higher (lower) use for other purposes as well. This interrelatedness, thus, defines the composition of the system.

We assume that, on average, organizations make directionally appropriate design decisions when confronted with a control problem, which implies that (again on average) they get the composition at least approximately right. Organizations, however, do not necessarily

optimize the use of the system to maximize effort direction effectiveness; instead, they will optimize their total set of control practices with regard to the overall performance of their organizations (Grabner and Moers, 2013; Masschelein and Moers, 2020).[5] While the problem of effort direction is important in conditions of low contractibility, it is only one control problem that organizations face (Bedford, 2020); for example, they also need to address problems of motivation and personal limitations. The solutions to these latter problems are not necessarily consistent with the solution needed to achieve effort direction effectiveness. Thus, organizations need to optimize a global control strategy to maximize overall performance that may require a trade-off among specific control problems. Because the relative importance of the various goals to achieve overall performance may differ across organizations, the extent of adoption and alignment of the system varies across organizations and organizations will differ in their effort direction effectiveness aspirations. Thus, a separate decision is for organizations to establish the intensity of use of the nondyadic control system. We argue that the more an organization emphasizes the nondyadic control system comprised of the three performance measurement uses, the better the effort direction problem will be solved (and the higher will be effort direction effectiveness).

In a high contractibility setting, in contrast, the ability to set clear and measurable goals provides the opportunity to contract upon these goals; also, there is generally alignment between the 'true objectives' of the organization and the associated performance measures, and organizational actors understand how to channel their efforts to achieve results (Baker, 2002; Gibbons, 1998; Hofstede, 1981). As a result, effort direction is less of a problem and the use of performance measures requires a less intricate approach to solve it; instead, organizations may use performance measures to solve other, more important control challenges that may emerge in a high contractibility setting.[6]

Based upon the previous reasoning, we formulate the following hypothesis:

H1. When contractibility is low, the combination of operational, exploratory, and incentive-oriented uses of performance information forms a nondyadic system that mitigates the effort direction problem. However, when contractibility is high, while operational, exploratory, and incentive-oriented uses of performance information may form a combination, it will not act as a system that mitigates the effort direction problem.

3. Methods and constructs

3.1. Institutional setting

Our study is set in the public sector in the Netherlands. This institutional setting is suitable for the purposes of our research project because problems of effort direction (the focus of our study) are prevalent in the public sector, as we will argue below.

Objectives in public sector organizations tend to be ambiguous (Dixit, 1997). Part of this ambiguity is deliberately created, as vague objectives provide politicians additional flexibility to react to changes in the political environment (Hofstede, 1981), to prevent budget cuts in 'pet projects' (De Bruijn, 2002), and to decrease the extent to which they can be held accountable by the general public (Bevan and Hood, 2006). But ambiguity is also related to the fact that public sector organizations are typically answerable to several different constituencies with different, possibly conflicting objectives (Dixit, 1997). For instance, a public organization may be formally accountable to the bureaucratic leadership of the ministry (or ministries) that fund(s) its operations, but it is quite common that there is also considerable influence of the minister him- or herself, parliament, various pressure groups, the media, et cetera. Against this background, it is often said that sector organizations have multiple principals (Burgess and Ratto, 2003; Dixit, 1997) that need to negotiate the organization's agenda in a political process, leading to multidimensional organizational objective functions that typically involve conflicting subgoals (Rainey and Jung, 2015). These conflicting subgoals give rise to thorny issues regarding the establishment of the relative weights of each performance dimension, ultimately resulting in the inability to define a 'summary measure' to capture overall performance (Tirole, 1994). Monitoring performance and delivering incentives is complex in these circumstances. Each principal will try to arrange for a positive coefficient on elements of performance that (s)he perceives as important and negative coefficients on the other dimensions (Dixit, 1997), and the aggregate marginal incentive coefficient for each performance measure will decrease in the number of principals as preferences differ among them (Burgess and Ratto, 2003). In addition, public sector organizations typically work in policy domains in which proven recipes for success do not exist. Joint efforts of multiple stakeholders, including external parties over which the organization has no control, are required; thus, public sector workers are unable to fully predict the results of their efforts or whether alternative actions are needed. As a result, performance measurement tends to be partial and noisy, and effort direction problems are particularly real in public sector organizations.

The Dutch public administration system belongs to the so-called 'Rechtsstaat tradition', a tradition that it shares with various countries in continental Europe (Bach et al., 2017). The Netherlands probably has the strongest tradition in the use of performance measurement in the public sector in continental Europe (Van Dooren et al., 2015). Therefore, it is commonly suggested that the Netherlands holds a midway position between the Anglo-Saxon public interest model on the one hand and the more legalistic and bureaucratic models of France and the Mediterranean states on the other (Pollitt, 2002; Van de Walle et al., 2004).

Public sector organizations in the Netherlands have considerable operational, administrative, and financial autonomy. Funding of public sector organizations is generally based upon a mixture of input funding based on general characteristics of the organization (e.g., size, service area) as well as reimbursements based upon outputs (e.g., volume of products and services); the mixture is different across sectors and over time. Less efficient public sector organizations need to find funding for potential deficits in their budget or need to cut back on service levels. The autonomy of public sector organizations in the Netherlands includes the adoption, implementation and use of management control systems (within budgetary and legal boundaries). As a result, management control systems (including performance measurement practices) vary substantially across public sector organizations, even within the same sector.

3.2. Empirical approach: measuring of and testing for nondyadic systems

Prior literature has examined complementarities between two control practices, modeling these as interactions. By extension, to examine the systemic use of three or more control practices (i.e., a nondyadic system), the default option is to include a three-way (or other higher-order) interaction in the empirical model. The problem with this approach, however, is that interaction terms compound measurement error, biasing against finding results. This is already an issue for two-way interactions, but adding additional, higher-order terms to the analysis exacerbates the problem (Hartmann and Moers, 1999).

If the control practices indeed form a system, the value of each

[5] In for-profit companies, economic profits can be defined as the overall performance that is optimized. In public sector organizations, overall performance is more difficult to determine and cannot easily be captured in a single summary measure (see Section 3.1).

[6] Our theory does not preclude the emergence of some form of combination of performance measurement uses when contractibility is high; instead, our theory contends that if such a combination emerges in that setting, it is not related to the control problem of effort direction. It may, however, be related to some other control problem, for instance, motivation or coordination.

individual practice is affected by all other practices in the set, either directly or indirectly. This interrelatedness manifests itself in shared variance, i.e., the part of the variance in the practices that they have in common with all other practices. Common factor analysis allows the identification of this shared variance, while singling out the variance in the practices that is not shared and that, consequently, cannot be part of the systemic use. In that process, the analysis also partials out measurement error, making the common factor approach more robust to type II errors than the interaction method. An additional benefit of the common factor approach is that it can also be applied to systems that comprise more than three practices, in which case an analysis based on higher-order interactions really becomes unwieldy.

Our argument, thus, is that if a set of practices forms a nondyadic system, the practices must have common variance. To be sure, we do not suggest that this shared variance must comprise a large proportion of total variance. Because the individual control practices may also play a role in mitigating control problems other than the specific control problem that requires the systemic response, there may be substantial amounts of variance that is not shared among the set of practices (Speklé et al., 2017).[7] Neither do we posit that shared variance necessarily means that the practices form a system, as there may be other explanations for this communality. It may, for instance, be the case that the set of practices share a common cause that impinges on the individual practices in the same way. This does not imply a system; for a systemic use, the practices much complement each other to achieve some specific outcome (Grabner and Moers, 2013; Hofmann and Van Lent, 2017; Masschelein and Moers, 2020). For this reason, it is important to control for contingency factors that may affect reliance on the collective practices before running the factor analysis. A straightforward way to do this is to regress each control practice on these contingency factors and subsequently use the regression residuals in the common factor analysis. This procedure should increase our confidence that the shared variance will in fact reflect complementarities among the practices, rather than some joint underlying cause.

Essentially, we model the system as a second-order superordinate construct (Edwards, 2001), with the practices (after controlling for other antecedents) as reflective indicators. The approach is visualized in Fig. 1. In this operationalization, a higher (lower) score on the second-order construct represents a higher (lower) intensity of use of the *combination* of performance measure uses. The next step, then, is to examine the correlation between the factor and the organization's performance in dealing with the focal control problem. Finding a result consistent with our theory further increases confidence that, in this case, the common variance is indicative of the existence of a system, i.e., that the individual practices support each other to help mitigate the theoretically identified control problem. It is important to note that we do not suggest that increasing reliance on the control system is 'better' or 'more optimal'; we only suggest that, on average, organizations that increase the reliance on the control system (i.e., have a higher score on the factor) will on average perform better in terms of solving the identified control problem (i.e., effort direction in a low contractibility setting) than those that do not.

In our approach, we combine a 'demand specification' with a 'performance specification' (Hofmann and Van Lent, 2017; Masschelein and Moers, 2020). The demand specification tests whether control practices are positively correlated with each other after controlling for environmental factors. It assumes that there is a 'sufficient number' of organizations that simultaneously choose the optimal level of the practices taking into account the interdependencies between the practices and the environment; the extreme version of this assumption is that all organizations make optimal choices (in which case there would be no link with performance). The performance specification tests whether the interdependencies among practices is positively correlated with performance. It assumes that there is a 'sufficient number' of organizations that deviate from the optimal level for the practices; the extreme version of this assumption is that all firms make random choices (in which case there would be no correlations between the practices; cf. Masschelein and Moers, 2020). In empirical research, neither assumption is likely to hold (Brynjolfsson and Milgrom, 2012; Hofmann and Van Lent, 2017). Following the recommendations of Hofmann and Van Lent (2017), we use both specifications. Thus, our analysis is based upon the assumption that, while there is a sufficient number of organizations that simultaneously choose the optimal level of performance management practices taking into account the interdependency and the organization's environment, there is also a sufficient number of organizations that deviate from the optimal level for the practices to detect performance differences.

This assumption is reasonable in our research setting. As already discussed in Section 2.4, effort direction effectiveness is an important concern, but not the only control problem that public sector organizations care about. Thus, their control choices will be influenced by effort direction considerations, but they will not necessarily seek to maximize this performance dimension as there may be other, competing issues that need to be addressed so that the extent of adoption of the system will vary across organizations. These differences can be exploited to estimate performance effects of different adoption decisions.

Our empirical strategy is consistent with the recommendations provided in theoretical papers by Grabner and Moers (2013) and Masschelein and Moers (2020) to test for interdependencies between control practices. Specifically, in a well-defined setting, we study a narrow control problem and a performance effect that is directly related to the control problem and the decision process at hand. This narrow problem is important enough to have a substantial effect on overall organizational performance, but the very specific focus helps us avoid some of the (potentially systematic) error that is associated with broader measures in cross-sectional research designs (Grabner and Moers, 2013). We control for other potential outcome effects in our measurement of effort direction effectiveness that may conflate performance on this specific dimension. We also argue why we expect variation in our focal performance variable for the organizations in our sample and why a performance specification is likely to lead to acceptable estimates (Grabner and Moers, 2013; Masschelein and Moers, 2020). Additionally, we provide evidence for the existence of complementarities between the control practices in a demand specification before examining performance effects, and control for appropriate contingency factors in this first step (Masschelein and Moers, 2020).

A characteristic of the factor analysis approach is that it does not provide information on the specific pattern of the complementarities at work. That is, if we find evidence of systemic effects, we cannot differentiate between complementarities that arise from the joint use of the full set of practices in the analysis and synergetic effects that are caused by overlapping (or interacting) pairwise complementarities. Our theory holds that each individual practice requires both other practices. If this theory is correct, this interrelatedness will show up in the factor analyses. But suppose that in reality, the relationships are different, such that two practices hold no direct relationship to each other, but are both affected by a third practice. In this case, we are still likely to find that the former practices correlate, and we will still find substantial communality in the factor analysis. Thus, the factor analyses approach cannot unequivocally demonstrate the existence of a particular pattern of interrelatedness. But then again, the approach quite convincingly disproves the existence of a nondyadic system if a common factor cannot be identified. Additionally, if a common factor emerges, this result indicates that the interrelatedness between the practices extends beyond discrete, separable dyadic complementarities and that a nondyadic

[7] For this reason, the usual standards to evaluate convergent validity of factor analysis outcomes do not apply when using them to operationalize a system of control practices. At this stage, we do not have a clear idea as to thresholds for factor loading, nor do we have any specific suggestions regarding acceptable minima for average variance extracted (AVE).

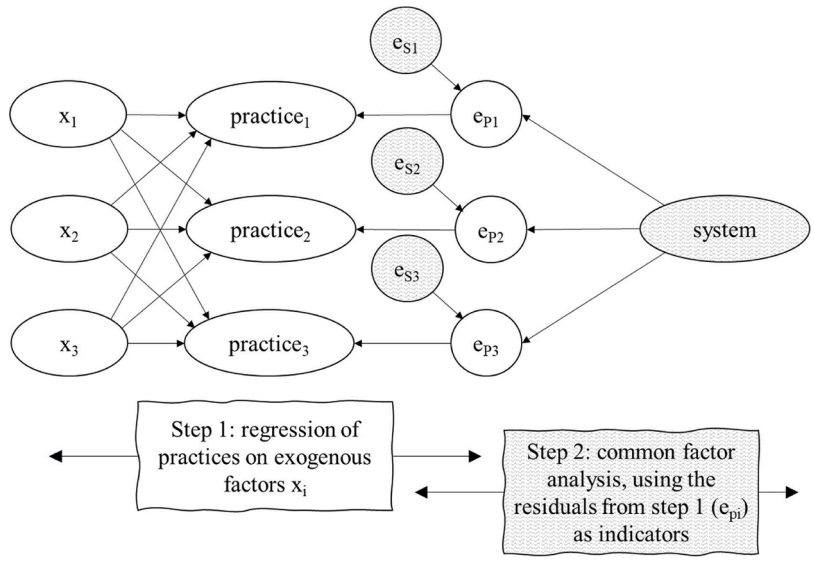


Fig. 1. Modelling Multi-Practice Systems.

perspective is required. Many empirical strategies need to settle for less.

3.3. Research design and sample

Our empirical study is based upon survey data from 162 organizational units in the public sector in the Netherlands. We use a convenience sample created through the personal networks of students in a Public Sector MBA class taught by one of the authors. Because we seek to test theory rather than to provide descriptive insights that generalize to broader populations, convenience sampling is justifiable for our purpose (Landers and Behrend, 2015; Speklé and Widener, 2018). Reliance on students' networks allows us to connect with and gain commitment from relevant respondents from different organizations within the public sector, helping us construct a diverse sample likely to contain the required variation in the variables of interest. We provided students with the survey instrument and an introduction letter explaining the broad purpose of the survey. To ensure that responses would be valuable for our research purposes, we gave very specific instructions to the students regarding the characteristics of the managers they could approach, and the units for which these managers are responsible. Participation in the survey process was entirely voluntary. We carefully instructed the students not to put any undue pressure on potential respondents to persuade them to participate. In addition, we made sure that students understood that there would be no repercussion if they would be unable to find potential respondents.

Respondents are managers with overall responsibility for the functioning of an organizational unit. We define such a unit as a relatively unified administrative entity within the larger organization, headed by a manager who has considerable authority over the set of tasks and processes of the unit (cf. Cavalluzzo and Ittner, 2004; Speklé and Verbeeten, 2014). The organizational unit is a suitable level of analysis for our project. Especially in larger organizations, performance measurement choices within the organization can be quite heterogeneous. Also, the degree of contractibility may differ significantly between the units that comprise the organization. This potential variety is less problematic at the unit level, where activities and control practices are usually much more homogenous. All observations in our dataset come from units responsible for activities that are typical of the public sector; we exclude more generic support units such as finance or human resource management from our analysis. We also exclude small units (number of employees <5 FTE) and units with newly appointed managers, i.e., managers with tenure of less than one year.

Our data show that respondents have been with their organization for 14 years on average (median 12 years) and have an average of almost four years of experience in their current function (median three years), with a minimum of one year. Given these employment histories, respondents can be trusted to be knowledgeable of the performance measurement practices in their organizational units, and the environment in which they are set. Table 1 provides descriptive detail about the units and the organizations to which they belong.

Even though we rely on single-source data, common method bias (CMB) does not pose a serious problem for our study. That is, if CMB is present, it will work against finding support for our hypothesis. A key element of the analysis is that we test for associations between a set of choices regarding the use of performance information and the organization's effectiveness in providing direction to its employees, predicting that we will find a significant association in some circumstances (i.e., low contractibility), but not in others (i.e., high contractibility). In the presence of CMB, we are likely to either find a relationship in both contractibility conditions, or none at all. Both outcomes would refute our theory. In addition, given the way in which we operationalize the effectiveness metric (see Section 3.3), a substantial level of CMB would leave too little variance in that metric, attenuating any associations between the metric and its antecedents. We nevertheless took care to reduce the potential for CMB, and following Podsakoff et al. (2003) we (1) use different response formats (e.g., 5-point scales, 7-point scales, point assignment questions), (2) pre-tested the survey instrument to

Table 1
Sample descriptives.

	Range	Mean	Standard deviation
Size organization (in FTE)	40–68,000	3,121	7,380
Size unit (in FTE)	6 – 1,500	118	263
		Number	**%**
Central government		47	29
Decentral government		63	39
Other public sector organizations		52	32
TOTAL		162	100

ensure clarity of wording,[8] (3) provided explicit assurance to respondents that there are no correct or incorrect answers to our questions, and (4) ensured full anonymity. Furthermore, we run a Harman's test using confirmatory factor analysis (Craighead et al., 2011). The (unreported) results show that a theory-based multifactor model fits the data considerably better than the alternative one-factor model. In sum, we conclude that it is unlikely that our inferences will be driven by CMB.

3.4. Construct measurement

To measure the constructs of interest in this study, we rely on instruments from prior literature,[9] albeit with some modification to the wording and set-up of the questions to ensure a proper fit with respondents' organizational contexts and their frame of reference. All multi-item first-order constructs are measured reflectively (Jarvis et al., 2003). As an initial step to explore the dimensionality in the data, we perform an exploratory factor analysis (EFA), using principal axis factoring with oblique rotation. To substantiate the EFA results, we apply confirmatory factor analysis in AMOS, including all multi-item first-order constructs. We find a well-fitting model, with CIMIN/DF = 1.439, TLI = 0.916, NNCP = 0.927, and RMSEA = 0.052.[10]

We examine indicator loadings[11] and average variance extracted (AVE) to assess convergent validity, cross loadings and the square root of AVE in relation with bivariate correlations to evaluate discriminant validity, and Cronbach's alpha to evaluate scale reliability. In addition, we examine criterion validity for all exogenous constructs in the analysis. All this is consistent with contemporary research standards in management accounting and control (Bedford and Speklé, 2018). Most AVE scores meet the conventional 0.5 benchmark (Fornell and Larcker, 1981; Henseler et al., 2009), except for two of the three dimensions of contractibility. These constructs, however, are subsequently combined in the formatively specified second-order construct that we use to measure contractibility (see below). A further analysis (not tabulated) shows that the full set of first-order constructs that ultimately end up in this second-order construct picks up 51 % of the variance in the underlying questionnaire items. For each reflectively measured construct, the square root of AVE is higher than the bivariate correlations between that construct and all other exogenous constructs in the analysis (Fornell and Larcker, 1981; details reported in Table 2), providing evidence of discriminant validity. Cronbach's alphas for all constructs meet or approach the rule of thumb of 0.7 (Hair et al., 1998).

To examine our hypothesis, we must run some of our analyses in two subsamples, i.e., a high and a low contractibility subsample. To test whether the factor structure of the first-order constructs is the same in these subsamples, we use a chi square difference test, comparing model fit between an unconstrained model and a model in which we constrain factor weights to be equal across both subsamples. The difference is insignificant ($p = 0.361$), providing evidence of measurement invariance and allowing us to report on construct measurement at the level of the full sample rather than the individual subsamples.

We calculate construct scores by taking the mean of the underlying item scores.[12] Table 2 reports descriptive statistics, including validity and reliability statistics if appropriate, while Appendix A reproduces the relevant parts of the survey and provides information on factor analyses outcomes and associated item selection decisions. Table 2 also reports the means of the construct scores separately for the two subsamples, alongside with information on the significance of the differences between the means. To avoid repetition, interesting differences will be discussed in conjunction with the examination of bivariate correlations in Section 4 and Table 3.

3.4.1. Effort direction effectiveness

Management control practices can serve a variety of different purposes, including motivation, decision support, coordination, empowerment, and effort direction (Bedford et al., 2016; Franco-Santos et al., 2007; Franco-Santos et al., 2012; Hansen and Van der Stede, 2004; Henri, 2006; Simons, 1990). Based on a relatively comprehensive listing of these purposes from Kruis (2008), we asked respondents to indicate how well the organization is performing with respect to five purposes: (1) motivating employees, (2) supporting decisions, (3) providing direction to employees' efforts, (4) coordination of effort, and (5) employee autonomy.

The effectiveness of control practices for the various purposes are likely to be interrelated. For instance, increasing employees' autonomy (empowerment) will often be associated with increased motivation. Moreover, apart from this real interrelatedness, scores could be affected by halo-effects, which occurs when there is spillover from a general judgment to a more specific one (Speklé and Widener, 2018). As we are interested solely in the contribution of a nondyadic system of performance measure use to solve the effort direction problem, we need to focus exclusively on the variance that is uniquely associated with effort direction effectiveness. To obtain such a clean measure, we regress effort direction success (#3 above) on the four remaining purposes. We use the residual of this regression as our metric for effort direction effectiveness.[13]

3.4.2. Use of performance measurement information

To measure the various uses of performance measurement information, we rely on a series of 14 questionnaire items based on Henri (2006); Hansen and Van der Stede (2004); Cavalluzzo and Ittner (2004); Speklé and Verbeeten (2014), and Verbeeten and Speklé (2015). [14] These items reflect operational use (e.g., operational planning and control, monitoring results, comparing outcomes to expectations), incentive-oriented use (e.g., performance assessment of individual employees, career decisions and bonuses for individual employees), and exploratory use (e.g., communicate goals, priorities and/or points of attention to employees, evaluate the validity of objectives and policy assumptions, assess the adequacy of policies and managerial decisions). In an exploratory factor analysis including all 14 items, each item loads on its expected construct, resulting in three multi-item variables representing the various performance information uses we distinguish theoretically.

To assess criterion validity of these three constructs, we examine

[8] We subjected the draft questionnaire to an academic pre-test by three experienced survey researchers, followed by an expert review by seven managers from public sector and not-for-profit organizations. This led to some minor adjustments in the wording of the questions.

[9] The only exception is the annual plan construct (see below), which is newly developed.

[10] In the selection of fit statistics, we follow Sharma et al. (2005). We apply common cutoff values to evaluate fit, i.e., CMIN/DF $\leq$ 2; TLI $\geq$ 0.90; NNCP $\geq$ 0.90; and RMSEA $\leq$ 0.08.

[11] We use 0.5 as the item retention threshold. On two occasions, however, we decided to retain an item with a factor loading of less than 0.5 (but larger than 0.4) to maintain correspondence with the original scale.

[12] We have chosen this approach to reduce the sensitivity of our measures to individual sample characteristics, allowing easier replication of our study in future work. Our substantive results are not affected by this choice, and our inferences remain the same if we use indicator weights to calculate construct scores.

[13] Note that because this residual includes information from a total of five different survey questions, our metric for effort direction effectiveness is *not* a single-item measure.

[14] The questionnaire also includes two measures that relate to accountability and legitimization (the use performance information for accountability to internal stakeholders and to external stakeholders; e.g, Henri, 2006; Cavalluzzo and Ittner, 2004). These have been excluded from the present analysis because this particular use is not part of our theory.

Table 2
Descriptive statistics.

	Range	Mean	Standard deviation	Mean low contractibility subsample	Mean high contractibility subsample	Significance difference means	Cronbach's alpha	AVE
Effort direction effectiveness	−2.89-2.55	0.000	.891	−.020	.018	ns	–	–
Use of performance information								
Operational use	1.00−5.00	3.298	.923	2.984	3.576	***	.861	.627
Incentive-oriented use	1.00−4.33	2.337	.938	2.013	2.624	***	.811	.608
Exploratory use	1.00−5.00	2.947	.877	2.673	3.188	***	.897	.599
Contractibility	−2.87−1.97	0.000	1.000	−.894	.790	***	–	–
Clarity of goals	1.00−5.00	3.710	.825	3.259	4.109	***	.829	.624
Measurability of output & goals	1.33−5.00	3.163	.788	2.614	3.647	***	.683	.436
Knowledge transformation process	1.25−4.75	3.172	.722	2.638	3.644	***	.732	.411
Control variables								
Data limitations	1.00−5.00	2.872	.841	3.000	2.759	*	.795	.517
Annual plan	1.00−5.00	3.268	.978	3.113	3.405	*	.869	.580
Politicization	0.00−50.00	13.690	10.774	16.18	11.48	***	–	–
Size unit (ln)	1.79−7.31	3.823	1.193	3.820	3.826	ns	–	–
Size organization (ln)	3.69−11.13	6.943	1.403	7.082	6.820	ns	–	–

ns = not significant; *p < 0.10, **p < 0.05, ***p < 0.01 (two-tailed).

Table 3
Bivariate correlations and square roots of AVE.

		1	2	3	4	5	6	7	8	9	10	11	12
1:	Effort direction effectiveness	–											
2:	Operational use	.088	.792										
3:	Incentive-oriented use	.063	.571***	.780									
4:	Exploratory use	.140*	.719***	.499***	.774								
5:	Contractibility	.038	.371***	.319***	.381***	–							
6:	Data limitations	.042	−.119	−.191**	−.017	−.122	.719						
7:	Annual plan	−.001	.180**	.061	.205***	.159**	.008	.762					
8:	Politicization	−.054	−.230***	−.164**	−.132*	−.231***	−.121	−.032	–				
9:	Central government	.181**	−.033	−.056	−.041	.043	−.113	−.005	.097	–			
10:	Decentral government	−.145*	−.145*	−.040	−.055	−.043	.054	.048	.280***	−.510***	–		
11:	Size unit (ln)	.109	.173**	.056	.116	.048	.113	.080	−.104	.131*	−.083	–	
12:	Size organization (ln)	.190**	.017	−.010	.164**	−.033	.181**	.002	−.067	.397***	−313***	.385***	–

Off-diagonal elements are Pearson correlations, diagonal numbers (underlined) are square roots of AVE.
*p < 0.10, **p < 0.05, ***p < 0.01 (two-tailed).

their associations with information on their antecedents or consequences. The use of performance information for operational purposes should help organizations to gain a clearer insight in their current functioning. In support of our measure, we find a rather strong and significant correlation between operational use and a survey item that picks up on the contribution of the performance measurement system to such insight ($r = 0.649$; $p < 0.01$). A more intense use of performance information for exploratory purposes aims to improve the organization's ability to respond to changes and unforeseen events, and to support the organization's ability to purposefully work on change and policy renewal. This is in fact what we observe in the data: the exploratory use measure is positively and significantly associated with scores on the survey items addressing responsiveness to changing conditions and the ability to manage change ($r = 0.550$ and 0.620, respectively; both p-values < 0.01). Finally, organizations with a stronger emphasis on an incentive-oriented use of performance information are more likely to value result-oriented individuals. Consistent with this, we find a correlation between incentive-oriented use and the appreciation of result-orientation as a trait of employees ($r = 0.129$; $p = 0.10$).

3.4.3. *Contractibility*

In Section 2, we defined contractibility as the degree to which desired contributions to the organization can be specified in advance. This depends on goal clarity, output measurability in terms of goal achievement, and knowledge of the transformation process. In line with this conceptualization and following Speklé and Verbeeten (2014), we model contractibility as a multidimensional second-order aggregate construct (Edwards, 2001) defined by three formative indicators corresponding to the dimensions of the construct. These formative indicators, in turn, are reflective constructs which we measure using previously developed instruments from the literature. The questionnaire items relating to clarity of goals and goal-consistent output measurability are based on Verbeeten (2008). Knowledge of the transformation process is measured with a slightly adapted version of the instrument described and analyzed in Whithey et al. (1983). We compute our contractibility proxy by constructing a linear composite of the three indicators, using the indicator weights to calculate the composite's scores. All three indicator weights are significant, and multicollinearity is not an issue with variance inflation factors ranging from 1.11 to 1.37 (Bedford and Speklé, 2018; Petter et al., 2007).

To establish criterion validity of the contractibility measure, we examine the association between this measure and additional survey information on the nature of the unit's responsibilities and tasks. We asked respondents to distribute a total of 100 points over three different descriptions of possible activities, highlighting policy development, policy implementation, and service provision to external clients as the main public sector activities. Public policy development is considerably more abstract and open-ended than policy implementation or the provision of specific services, and we expect units that are heavily engaged in policy development to report the lowest levels of contractibility.

Consistent with this expectation and in support of the validity of the contractibility measure, we find a significant negative association between policy development involvement and contractibility ($r = -0.445$; $p < 0.01$).

3.4.4. Control variables and variables for sensitivity analyses

As discussed in Section 3.1, we control for several contingency variables that may relate to the demand for and use of performance information. Apart from contractibility, we control for data limitations, the reliance on annual plans, the degree of 'politicization' of the unit's functioning, institutional context, and size. The reasoning underlying this selection of variables will be explained in Section 4 below; at this point, we confine ourselves to their operationalization.

To capture potential differences in the institutional context, we use information on the type of organization (central government, decentral government, and other public sector organizations) and we measure unit and organizational size as the natural log of the number of employees. The items to measure data limitations are taken from Cavalluzzo and Ittner (2004) and address difficulties in obtaining timely, valid or reliable data, data collection costs, and the inability of existing information systems to provide the needed data. The survey questions to capture the importance of the annual plan are self-developed. Factor analysis confirms unidimensionality of both constructs. To proxy for the degree of politicization, we use a single-item measure indicating the extent to which the organization values individuals' political savviness, relative to other competencies and skills.

In support of criterion validity of the data limitation construct, we find a negative association between that construct and a four-item construct capturing routine access to various types of quantitative information ($r = -0.409$; $p < 0.01$). One would expect that the importance of the annual plan also affects performance evaluation. Consistent with this expectation, we find a positive and significant correlation between the organizational status of the annual plan and the relative weight of realized results versus plan in formal performance assessments ($r = 0.291$; $p < 0.01$). To corroborate criterion validity of politicization, we assess its association with a survey question asking about the extent to which the unit's goals are affected by political developments. We expect and find a positive correlation ($r = 0.188$; $p < 0.05$).

In a robustness analysis reported in Section 4 below, we will seek additional assurance that the patterns we observe in the use of performance information are truly driven by effort direction problems (as per our theory) rather than some other purpose of performance measurement. For this analysis, we calculate a series of alternative effectiveness measures, i.e., motivational effectiveness, decision support effectiveness, coordination effectiveness, and empowerment effectiveness. Like our metric for effort direction effectiveness, we calculate these measures as the residual of a regression of the unit's performance in achieving that particular objective on the four other control purposes on which we have information.

4. Results

Table 3 presents the correlation matrix and reports the square root of AVE for each reflectively measured construct.

Effort direction effectiveness is associated with exploratory use of performance information, but not with the other uses of performance information nor with the use of qualitative annual plans. Interestingly, effort direction effectiveness and contractibility appear to be uncorrelated, which may be taken as a very preliminary indication that, on average, organizations have found ways to address the effort direction problems that come with low contractibility. Contractibility itself is positively associated with all three uses of performance information, as well as with reliance on the qualitative information from the annual plan. This insight was already apparent from the subsample descriptive statistics in Table 3 that show that the mean use of performance data and annual plan information is significantly lower in the low contractibility subsample than in the group of respondents reporting high levels of contractibility. This finding is consistent with the notion that public sector organizations characterized by low contractibility may emphasize other types of control (Hofstede, 1981; Ouchi, 1979).

Like previous studies in the public sector (Speklé and Verbeeten, 2014; Verbeeten and Speklé, 2015) as well as studies in the private sector (Hansen and Van der Stede, 2004), we find strong correlations between the three uses of performance information, ranging from 0.499 (between exploratory and incentive-oriented use) to 0.719 (between operational and exploratory use). An interesting finding is the negative correlation between politicization on the one hand, and contractibility and the uses of the performance measurement system on the other. This finding is consistent with Hofstede's argument that if organizational goals are ambiguous, control acquires a predominantly political flavor (Hofstede, 1981).

4.1. First-stage regressions (step one)

A key claim in our theory is that in conditions of low contractibility, the three uses of performance information are complements and the contribution of each individual practice to effort direction effectiveness depends on both other practices. If this claim is valid, the three practices must share (part of) their variance. However, shared variance may also be caused by other factors that affect the focal control practices in similar ways. Therefore, we first regress each control practice on a series of potential joint causes to control for such confounding factors (cf. Grabner and Moers, 2013; Masschelein and Moers, 2020). We control for the main effect of contractibility on the individual control practices. In addition, we take the broader informational setting in which the units operate into account, because this may provide additional clues (or create additional barriers) for employee effort direction decisions. Particularly, we control for data limitations and the reliance on so-called annual plans. The data limitations construct refers to technical problems with existing information systems and picks up on the difficulties people experience in getting timely access to relevant and reliable information. Earlier studies have found that such technical issues play an important role in the implementation and use of performance measurement systems (Cavalluzzo and Ittner, 2004). Annual plans specify actions, projects, and priorities for the upcoming period in qualitative terms. Although the adoption of these plans is sometimes voluntary and their actual use may differ across organizations, their implementation is often required by Dutch law or governance regulation. For this reason, we treat the reliance on an annual plan as an exogenous variable. The annual plan could serve as a substitute for quantitative performance information, perhaps dependent on differences in management styles (e.g., a more qualitative management style versus a 'numbers-based' approach; Speklé and Verbeeten, 2014). Alternatively, one could argue that the annual plans may provide a frame of reference for the interpretation of quantifiable performance information, facilitating and intensifying the use of such number-based information.

We also control for the politicization of the units, i.e., the extent to which political considerations play a role in the functioning of the units. Prior research has documented that political processes affect the meaning attached to performance information (e.g., Lewis, 2015), which may influence the reliance on and use of such information. Finally, we include information on the institutional and organizational context, i.e., the type of organization (central government, decentral government, and other public sector organizations), the size of the unit, and the size of the organization. Organizational type may be important as different branches of the public sector are subject to different regulatory regimes and may face different demands from stakeholders, while unit and organizational size may influence the reliance on formal control practices, like the use of performance measurement information, and may affect the ability to invest in such practices (Chenhall, 2003).

Table 4 presents the results of the first-stage regressions. We find that each of the three performance measurement uses is affected by

Table 4
First-stage regressions.

	Operational Use	Incentive-oriented use	Exploratory use
Contractibility	.307***	.275***	.368***
Data limitations	−.166	−.195**	−.033
Annual plan	.124*	.012	.143*
Politicization	−.088	−.094	−.009
Central government	−.156*	−.125	−.189**
Decentral government	−.178*	−.037	−.072
Size unit (ln)	.165**	.050	.025
Size organization (ln)	−.018	.046	.223**
F-value (p)	5.686 (.000)	3.420 (.001)	5.404 (.000)
R^2	.229	.152	.220
adj. R^2	.189	.107	.180

*p < 0.10, **p < 0.05, ***p < 0.01; two-tailed.

contractibility. This is expected since contractibility problems negatively affect the quality of performance measures, limiting their usefulness. The results further indicate that technical issues with the information system hinder the use of performance information for incentive purposes, but not for operational and exploratory objectives. This suggests that an incentive-oriented use puts more stringent demands on information quality than the other two uses. We find no effects for politicization, but the evidence suggests that the reliance on annual plans intensifies the use of performance information for operational and exploratory purposes. This is consistent with the suggestion that such plans facilitate rather than supplant the use of quantitative performance data. Finally, we find some sector differences and a few size effects.

4.2. Factor analysis and effectiveness correlations (steps two and three)

Using the residuals of the first-stage regressions reported in Table 4 as indicators, we perform a factor analysis to estimate the remaining overlap in the three performance measurement uses following the removal of the effects of joint determinants. To test our hypothesis, we split our sample at the mean of contractibility into a high and a low contractibility sample.[15] The low group contains 76 organizational units (46.9 percent); the high group holds 86 observations (53.1 %). Applying principal axis factoring, we find a one-factor solution in both subsamples. We use factor scores (indicator weights) to calculate the combination of performance measurement uses found in the low and high contractibility conditions.[16] We use the term 'PM factor' to refer to these constructs. Table 5 reports the factor analysis outcomes.

The factor solutions differ substantially across the subsamples. Although we had no prior expectations as to the relative weightings among the three uses, we find that in the low contractibility condition, all three uses contribute to the score on the PM factor, with indicator weights ranging from 0.199 to 0.436. Factor loadings range from 0.688 to 0.842, indicating that the latent construct explains 47 % or more of the variance in each individual use. The results are quite different in the high contractibility condition. Even though we find a single factor solution, this PM factor is now effectually 'dominated' by operational use as one of the three uses of performance information. That is, while the indicator weight of operational use is 0.839, the other two uses add very little to the construct, with indicator weights of only 0.052 (incentive-oriented use) and 0.130 (exploratory use). Moreover, the factor loadings range from 0.454 to 0.946, indicating that the latent construct explains about 90 % of the variance in operational use but only about 21 % of the variance in incentive-oriented use.

Table 5
Factor analysis and correlation effectiveness.

	Low contractibility (n = 76)		*High contractibility (n = 86)*	
	Factor loading	Indicator weight	Factor loading	Indicator weight
Operational use[i]	.842	.436	.946	.839
Incentive-oriented use[i]	.688	.199	.454	.052
Exploratory use[i]	.834	.413	.705	.130
Average variance extracted	.626		.533	
Correlation with effort direction effectiveness[ii]	.256 (p = .026)		−.033 (p = .765)	

Reported p-values are two-tailed.
[i] Based on the residuals of the first-stage regressions (see Table 4).
[ii] We test the difference between the two correlations using Fisher's *r* to *z* transformation. Our test results indicate that the difference between the correlations is significant (p = 0.066; two-tailed).

Our theory implies that in conditions of low contractibility, the PM factor represents a system of performance measurement uses that helps mitigate the effort direction problem. Higher (lower) scores on this factor corresponds to a higher (lower) intensity of use of the combination of performance measurement uses. This is visualized in Fig. 2.[17] The next step, then, is to examine the correlation between the PM factor and the organization's performance in dealing with the effort direction problem.

Consistent with our hypothesis, we find that the PM factor is correlated with the outcome measure in the low contractibility subsample (r = 0.256; p < 0.05, two-tailed), but not in high contractibility (r = −0.033, n.s.). We use Fisher's *r* to *z* transformation and compare the test statistic (see Millsap et al., 1990); our test results suggest that the difference between the correlations is significant (p < 0.10, two-tailed). This finding is consistent with our theory: the PM factor is correlated with the predicted specific control outcome in the condition we specified in advance, and only in that condition. Because a performance specification is especially vulnerable to Type I error, we follow the suggestion of Masschelein and Moers (2020) and apply bootstrapping to estimate the bias-corrected 95 % confidence interval; again, we find support for a significant positive correlation in the low contractibility condition. We proceed to explore potential associations with alternative control purposes (motivation, decision support, coordination, and autonomy support). In the low contractibility condition, we find none. This suggests that the PM factor in this condition is in fact driven by effort direction concerns and not by some other purpose of management control that just happens to be associated with effort direction effectiveness. Interestingly, but on a side note, we find that the PM factor in the high contractibility setting is associated with coordination effectiveness (r = 0.312; p < 0.01). This could imply that the PM factor in this setting also represents a system, but that it is used to solve another control problem that may be more prevalent in that setting.[18] We leave this suggestion to future research. In sum, our results are consistent with our

[15] The analysis in subsamples is based on our theoretical motivation that a systemic use of performance measures is associated with effort direction effectiveness in a low contractibility setting yet not in a high contractibility setting. Therefore, we are interested in how the control practices combine in that particular condition and we do not want the estimate of their interrelatedness to be affected by what transpires in the other setting.

[16] We use the regression scores option in SPSS to calculate the factor weights. Since there is only one factor, the Bartlett scores option in SPSS returns a perfectly correlated factor to that which is obtained using the regression scores option. Thus, the choice of methods to calculate factor scores does not influence our results.

[17] We reiterate the point made earlier that a higher score of the PM factor represents a more intense use of the combination of practices, not a better or 'more optimal' system.

[18] As the latent construct explains about 90% of the variance in operational use in the high contractibility setting, the focus may indeed be on coordination and planning of operations.

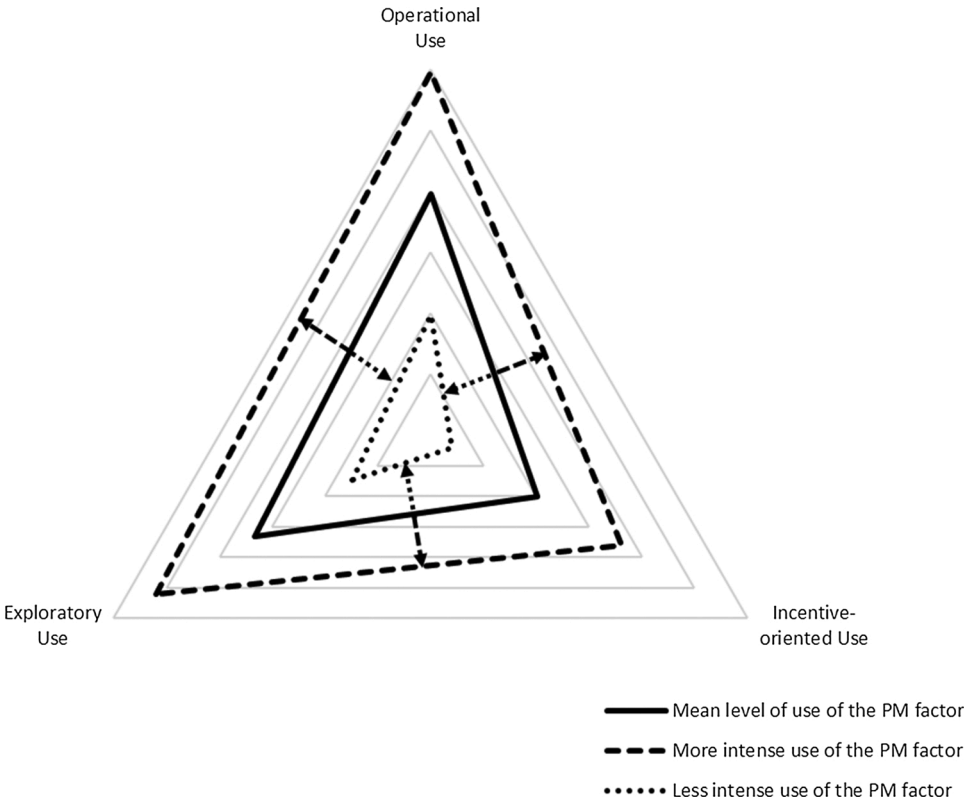


Fig. 2. Varying Intensity of the Use of the PM Factor.
Note: As the figure above shows, the ratio of reliance on the three performance measurement is held constant; that is, the composition of the system is held constant. As an organization moves up or down on the factor score line (i.e., varies the intensity of the PM factor or system of performance measurement uses), the organization will achieve more or less effort direction effectiveness.

hypothesis that a system comprised of three complementary performance measurement uses is associated with effort direction effectiveness when contractibility is low.

4.3. Exploring alternative explanations

In the base-line analyses reported above, we split the sample at the mean of contractibility. This may be considered an arbitrary choice that may have affected the outcomes of the analyses. Therefore, we rerun the analyses in redefined subsamples, now split at the median. This means that five observations that were in the high contractibility sample are now moved to the low contractibility group. The results (untabulated) remain the same and our conclusions appear to be robust to these subsample definition choices.

As stated in Section 3.3, we calculated the scores on all reflectively measured first-order constructs using the mean of the associated item scores.[19] This choice is defensible from a replication perspective but implies that the resulting constructs contain both unique and error variance which may have affected our findings. To explore this possibility, we reran the full set of analyses using construct scores based on indicator weights instead. An additional consequence of this exercise is that one observation that was in the low contractibility group now moves to the high condition, while three other observations transfer from the high to the low contractibility subsample. In these new (untabulated) analyses, our substantive findings are qualitatively similar to the initial examination, and our inferences are not affected in any way.

[19] Please note that this only relates to the first-order constructs. The second-order superordinate construct that measures the PM system has, of course, been calculated using factor weights.

5. Conclusion

In this paper, we examine whether three performance measurement uses combine in a nondyadic system. Providing effort direction using performance measures is difficult in settings where goals tend to be ambiguous, outcomes are difficult to measure, and knowledge of the transformation process is imperfect at best –like the public sector (Burgess and Ratto, 2003; Hofstede, 1981; Tirole, 1994). In such a low contractibility setting, organizations need to combine operational, incentive-oriented, and exploratory uses of performance information in a nondyadic way to address the effort direction problem. We argue that this is because each performance measurement use has limitations and potential unintended consequences, but when used in combination with the other two uses, its benefits towards achieving effort direction effectiveness are realized. That is, the positive effect of one use depends on the other two uses. However, because our empirical method is limited to providing evidence about the existence of a system and does not provide evidence on the internal structure of such system, we hypothesize only that the three performance measurement uses combine in a single system to mitigate the problem of effort direction when contractibility is low. Our empirical evidence supports these arguments, documenting that in low (high) contractibility, the three different uses of performance measurement form a nondyadic system (a combination) that contributes (is unrelated) to effort direction effectiveness.

Our study illustrates an empirical technique to identify, measure, and test a nondyadic system. Building on the criteria put forth by Grabner and Moers (2013), we suggest that there are two cumulative requirements necessary to claim the existence of such a nondyadic system. First, a factor analysis of the residuals of the control practices, after controlling for a well-argued set of contingency factors, must result in a one-factor solution; this is a test of a necessary yet not sufficient condition. Second, there must be evidence that the combination of control practices helps solve the theorized control problem. In our empirical approach, we follow these recommendations, taking into account the suggestion from Hofmann and Van Lent (2017) and Masschelein and

Moers (2020) to provide both a demand and performance specification for the systemic use of performance measurement practices in public sector organizations. We test the correlation of the systemic use of performance measures in a research setting where variation likely exists in performance (i.e., effort direction effectiveness) across organizations, controlling for performance effects related to other control problems.

An appealing feature of our approach is its ability to include a relatively large number of control practices. In the current paper, we address the use of three performance measurement practices, but the number of control practices could be expanded if theory suggests a more complex system. Note in this context that, apart from several studies that use cluster analysis techniques to map control patterns (e.g., Bedford and Malmi, 2015; Kruis et al., 2016), most prior studies examining a control system confine themselves to a limited number of practices, usually addressing discrete pairwise interrelations (Bedford, 2020; Hofmann and Van Lent, 2017; Masschelein and Moers, 2020). This dyadic emphasis may sometimes be predicated on theoretical considerations, but we suspect that it may also originate from the technical difficulties that were hitherto associated with the analysis of complementary relations between more than two control practices (Speklé and Widener, 2020). The factor analysis approach we apply in this study may offer a way forward. We believe that our approach could apply quite generally, and we hope that our operationalization of a system comprised of three performance measurement uses opens many interesting new avenues of research focusing on other nondyadic combinations of control practices. Even though this approach does not provide evidence on the specific pattern of the complementarities at work and cannot distinguish between, for instance, the synergetic effects from a three-way interrelation between practices and multiple interconnected pairwise complementarities, it does pick up on interrelations beyond singular pairwise interactions, providing a valuable clue to researchers as to whether more complex relationships between control practices should be considered. At this point, it should be noted that even though the pattern is of interest theoretically, it does not matter much from a practical point of view; whatever the pattern, the practices should still be used together, and if the organization increases its reliance on one practice, it is best advised also to increase its reliance on the other practices.

Our study reveals several alternative research questions. One obvious suggestion would be to explore the private sector response to effort direction problems. Even though we tested our theory in a public sector context, contractibility challenges are not unique to that particular sector, and the theory we developed is not restricted to public sector organizations. Therefore, our findings may well carry over to the less programmable parts of private sector activity, such as R&D activities and creative professional organizations. Future research can also consider additional theoretically motivated linkages amongst specific contingency factors, the associated control problems and subsequent control choices in organizations, and evaluate how they affect the performance of the organization on the specific dimension that would be affected by these control practices.

This study also contributes to the current debate about the role of performance measurement practices in organizations. Previous literature has identified the determinants and consequences of specific uses of performance measurement information (e.g., Franco-Santos et al., 2007; Hansen and Van der Stede, 2004; Henri, 2006). Our study provides additional insights into how these uses combine in a system to contribute to organizational effectiveness in conditions of low contractibility, an organizational circumstance that has been identified as a key determinant of management accounting systems (Chenhall, 2003).

Finally, we contribute to the current debate about the role of performance management in public sector organizations. New Public Management has been guiding public sector reforms for over 25 years (Hood, 1995; OECD, 1994), emphasizing the role of performance management. Like the broader literature, previous studies in this specialized field have focused on the main effects of performance management practices (e.g., Burgess et al., 2017; Van der Kolk et al., 2019), but have ignored the interrelatedness between these practices. We demonstrate that there may be complementarities among the performance measurement practices, both theoretically and empirically, thus providing a more nuanced understanding of the role of performance measurement practices in public sector organizations.

This study is subject to several potential limitations. Our evidence is based on a convenience sample with Dutch respondents only. A single country sample provides indirect control over potentially confounding effects of cultural differences and institutional factors but limits the generalizability of the findings. In addition, since our observations relate to organizational sub-units, we are unable to assess the representativeness of our sample. However, our main concern is not with generating insights that apply to a broader population, but with testing theory. For that purpose, representativeness of the sample is less of an issue than sample relevance (Speklé and Widener, 2018). As we argued in Section 3.1, the public sector in the Netherlands displays a high level of administrative autonomy regarding the adoption, implementation and use of management control practices, leading to considerable variety across organizations and providing for an interesting setting for our study. Furthermore, our sample is heterogeneous, including respondents from different branches of the public sector and covering organizations and units of different sizes and with different tasks and responsibilities, and seems at least informative of experience in the Netherlands. Nevertheless, further research is required to assess whether the results can be replicated in a broader public sector setting and in countries with other institutional settings.

Another potential limitation of our study is in the calculation of the residuals of the control practices that enter the factor analyses to examine the interrelatedness between control practices. Besides including contractibility in our first-stage regressions, we control for size effects and type of organization as the 'usual suspects' (consistent with, for instance, Abernethy et al., 2015) and include controls for the broader information setting and politicization of the units. It could nevertheless be argued that a richer set of potential drivers of the demand for the control practices should have been included. Our data, however, did not allow such broader examination. We do, though, control for alternative control objectives when calculating our measure of effort direction effectiveness. These specification choices and checks should help alleviate omitted variable concerns. Finally, as discussed earlier, our empirical method cannot provide evidence on the specific patterns of the complementarities found in the structure of the system.

Despite the caveats, our study has various practical implications for the design and use of performance measurement systems. Recent years have seen numerous calls to strengthen incentives within organizations, 'bringing the market inside the firm or organization' (Roberts, 2010); the supposed advantages are that people will work harder and smarter if they are rewarded for results. However, strong incentives may be problematic, especially within complex organizations with difficult to measure results; the problem is that people respond just as strongly to badly designed incentives as they do to well-structured ones (Roberts, 2010: 125). In a low contractibility setting, complementary uses of performance measures are necessary to provide effort direction. This more nuanced view of performance measurement practices may be beneficial to support practical performance management decisions in low contractibility circumstances in settings like the public sector and R&D organizations.

Acknowledgements

We thank Theresa Libby (the editor) and two anonymous reviewers for their constructive support and many insightful comments. We thank Henri Dekker, Frank Hartmann, Peter Kroos, Victor Maas and Laurence van Lent for very productive discussions that helped us sharpen our thinking and reasoning. Previous versions of this paper have been presented at the 2018 AOS conference on Management Control as System or

Package, the 2019 Management Accounting Section Conference, and research seminars at the University of Groningen and Nyenrode Business University. We thank conference and seminar participants for their comments and suggestions, particularly Martin Artz and Paula van Veen-Dirks.

Appendix A. Survey instrument (abridged)

Factor loadings are reported in parentheses.

Contractibility

Please indicate the extent to which you agree with the following statements:

Clarity of goals

- The mission of your unit is formulated unequivocally (.843)
- The mission of your unit is actively propagated (.704)
- The objectives of your unit unambiguously related to your unit's mission (.816)

Measurability of output & goals

- Your unit's goals can well be translated in measurable results (.808)
- To get a good picture of your unit's performance, it suffices to look at the measurable results achieved (.557)
- The effect of your unit's efforts become visible within a year (.589)

Knowledge transformation process

- For your unit's activities, there is a logical sequence of steps to be taken (.697)
- Your unit's primary processes can only be performed effectively in one specific way (.608)
- Within the unit, it is entirely clear how to perform our tasks (.678)

Use of performance information

Please indicate the intensity with which you use the performance measurement system and the information it contains for the following purposes:

Operational use

- You use information from the performance measurement system for operational planning (allocation of capacity, task assignment) (.628)
- You use information from the performance measurement system to allocate budgets within your unit (.405)
- You use information from the performance measurement system to monitor operational activities (.772)
- You use information from the performance measurement system to monitor the performance of your unit (.873)
- You use information from the performance measurement system as a signal to intervene if performance falls short of expectations (.862)

Incentive-oriented use

- You use information from the performance measurement system to assess the performance of individual employees within your unit (.733)
- You use information from the performance measurement system for career decisions with regard to individual employees (.929)
- You use information from the performance measurement system to determine bonuses (.652)

Exploratory use

- You use information from the performance measurement system to communicate goals, priorities and / or points of attention to employees of your unit (.586)
- You use information from the performance measurement system to weigh different policy or implementation alternatives (.773)
- You use information from the performance measurement system to evaluate the validity of objectives and / or policy assumptions (.786)
- You use information from the performance measurement system to review or adjust the policy of the unit (.826)
- You use information from the performance measurement system to assess the adequacy of your policies and managerial decisions in hindsight (.834)
- You use information from the performance measurement system to get a better overview of the consequences of actions taken or choices made in hindsight (.813)

Effort direction effectiveness

In your experience, how does the control system perform on the following dimensions?

- Motivating and stimulating employees
- Supporting decisions
- Providing direction to employees' efforts
- Coordination of effort
- Employee autonomy

We measure effort direction effectiveness as the residual from a regression of the third item on the four other dimensions. Likewise, we calculate alternative effectiveness measures (i.e., motivational effectiveness, decision-making effectiveness etc.) by estimating the residual from the relevant item on the other dimensions.

Control variables

Data limitations

Please indicate the extent to which you agree with the following statements about information in your organization:

- It is very difficult to obtain valid and reliable information (.858)
- It is very difficult to obtain information in time (.820)
- The costs of collecting data are very high (.574)
- Existing IT-systems are unable to provide the necessary information (.573)

Annual plan

We are also interested in the possible role of annual plans. The term 'annual plan' refers to an activity plan, policy plan and/or implementation plan that describes actions, projects, and priorities for the upcoming period. In this question, we are explicitly not referring to quantified performance contracts. Such performance contracts are discussed elsewhere in the questionnaire. We ask for your opinion on the following statements. If there are no such plans within your unit, please check the '1' in the table.

- Your unit's efforts are based on an annual plan that has been approved by your organization's senior management (.754)
- The annual plan provides strong guidance for the efforts of your unit (.757)
- In case of deviation from the annual plan, explicit approval from senior management is required (.624)
- Progress of the annual plan's realization gets intensive attention from senior management (.842)

- Progress of the annual plan's realization is a frequent subject of consultation between you and senior management (.812)

Politicization

We are interested in the appreciation that exists within your organization for certain characteristics of people. To obtain an overall picture of this, we ask you to divide exactly 100 points in the table below, in such a way that the number of points per category indicates the relative valuation of that category. Of course, you can choose to assign all 100 points to a single category or to give certain categories 0 points, if that matches reality.

- Result orientation
- Feeling for political and/or administrative relations
- Subject-matter expertise
- Skills relating to process management
- Social and communication skills
- Respect for procedures
- Other properties than the above

We proxy for politicization using the score on the second item.

References

Abernethy, M.A., Dekker, H.C., Schulz, A.K.D., 2015. Are employee selection and incentive contracts complements or substitutes? J. Account. Res. 53, 633–668.
Argyris, C., 1977. Organizational learning and management information systems. Account. Organ. Soc. 2, 113–123.
Argyris, C., 1990. The dilemma of implementing controls: the case of managerial accounting. Account. Organ. Soc. 15, 503–511.
Bach, T., van Thiel, S., Hammerschmid, G., Steiner, R., 2017. Administrative tradition and management reforms: A comparison of agency chief executive accountability in four Continental Rechtsstaat countries. Public Manag. Rev. 19, 765–784.
Baker, G., 2002. Distortion and risk in optimal incentive contracts. J. Hum. Resour. 37, 728–751.
Bedford, D.S., 2015. Management control systems across different modes of innovation: implications for firm performance. Manag. Account. Res. 28, 12–30.
Bedford, D.S., 2020. Conceptual and empirical issues in understanding management control combinations. Account. Organ. Soc. 86, 101187.
Bedford, D.S., Malmi, T., 2015. Configurations of control: an exploratory analysis. Manag. Account. Res. 27, 2–26.
Bedford, D.S., Speklé, R.F., 2018. Construct validity in survey-based management accounting and control research. J. Manag. Account. Res. 30 (2), 23–58.
Bedford, D.S., Malmi, T., Sandelin, M., 2016. Management control effectiveness and strategy: an empirical analysis of packages and systems. Account. Organ. Soc. 51, 12–28.
Bevan, G., Hood, C., 2006. What's measured is what matters: targets and gaming in the English public health care system. Public Adm. 84 (3), 517–538.
Brynjolfsson, Erik, Milgrom, Paul, 2012. Complementarity in Organizations. In: Gibbons, Robert, Roberts, John (Eds.), The handbook of organizational economics. Princeton University Press, pp. 11–55.
Burchell, S., Clubb, C., Hopwood, A., Hughes, J., Nahapiet, J., 1980. The roles of accounting in organizations and society. Account. Organ. Soc. 5, 5–27.
Burgess, S., Ratto, M., 2003. The role of incentives in the public sector: issues and evidence. Oxford Rev. Econ. Policy 19, 285–299.
Burgess, S., Propper, C., Ratto, M., Tominey, E., 2017. Incentives in the public sector: evidence from a government agency. Econ. J. 127 (605), F117–F141.
Burney, L., Widener, S.K., 2007. Strategic performance measurement systems, job-relevant information, and managerial behavioral responses-role stress and performance. Behav. Res. Account. 19 (1), 43–69.
Cavalluzzo, K.S., Ittner, C.D., 2004. Implementing performance measurement innovations: evidence from government. Account. Organ. Soc. 29, 243–267.
Chenhall, R.H., 2003. Management control systems design within its organizational context: findings from contingency-based research and directions for the future. Account. Organ. Soc. 28, 127–168.
Chenhall, R.H., Morris, D., 1995. Organic decision and communication processes and management accounting systems in entrepreneurial and conservative business organizations. Omega 23, 485–497.
Craighead, C.W., Ketchen, D.J., Dunn, K.S., Hult, G.T.M., 2011. Addressing common method variance: guidelines for survey research on information technology, operations, and supply chain management. IEEE Trans. Eng. Manag. 58, 578–588.
Davila, A., Foster, G., Li, M., 2009. Reasons for management control systems adoption: insights from product development systems choice by early-stage entrepreneurial companies. Account. Organ. Soc. 34, 322–347.
De Bruijn, H., 2002. Performance measurement in the public sector: strategies to cope with the risks of performance measurement. Int. J. Public Sect. Manag. 15 (7), 578–594.
Dixit, A., 1997. Power of incentives in private versus public organizations. Am. Econ. Rev. 87 (2), 378–382.
Edwards, J.R., 2001. Multidimensional constructs in organizational behavior research: an integrative analytical framework. Organ. Res. Methods 4, 144–192.
Emmanuel, C., Otley, D., Merchant, K., 1990. Accounting for Management Control. Springer, Boston, MA.
Feltham, G.A., Xie, J., 1994. Performance measure congruity and diversity in multi-task principal/agent relations. Account. Rev. 69, 429–453.
Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. 18 (February), 39–50.
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. Int. J. Oper. Prod. Manage. 27, 784–801.
Franco-Santos, M., Lucianetti, L., Bourne, M., 2012. Contemporary performance measurement systems: a review of their consequences and a framework for research. Manag. Account. Res. 23 (2), 79–119.
Gibbons, R., 1998. Incentives in organizations. J. Econ. Perspect. 12 (4), 115–132.
Gibbons, R., Kaplan, R.S., 2015. Formal measures in informal management: can a balanced scorecard change a culture? Am. Econ. Rev. 105 (5), 447–451.
Grabner, I., 2014. Incentive system design in creativity-dependent firms. Account. Rev. 89, 1729–1750.
Grabner, I., Moers, F., 2013. Management control as a system or a package? Conceptual and empirical issues. Account. Organ. Soc. 38, 407–419.
Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis. Prentice Hall, Inc., Upper Saddle River, NJ.
Hansen, S.C., Van der Stede, W.A., 2004. Multiple facets of budgeting: an exploratory analysis. Manag. Account. Res. 15, 415–439.
Hartmann, F.G., Moers, F., 1999. Testing contingency hypotheses in budgetary research: an evaluation of the use of moderated regression analysis. Account. Organ. Soc. 24 (4), 291–315.
Henri, J.F., 2006. Organizational culture and performance measurement systems. Account. Organ. Soc. 31, 77–103.
Henseler, J., Ringle, C.M., Sinkovics, R.R., 2009. The use of partial least squares path modeling in international marketing. Adv. Int. Market. 20, 277–319.
Hofmann, C., van Lent, L., 2017. Organizational design and control choices. In: Hitt, M.A. (Ed.), The Oxford Handbook of Strategy Implementation.
Hofstede, G., 1981. Management control of public and not-for-profit activities. Account. Organ. Soc. 6, 193–211.
Holmström, B., 2017. Pay for performance and beyond. Am. Econ. Rev. 107 (7), 1753–1777.
Holmstrom, B., Milgrom, P., 1991. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. J. Law Econ. Organ. 7, 24–52.
Hood, C., 1995. The 'new public management' in the 1980s: variations on a theme. Account. Organ. Soc. 20, 93–109.
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. J. Consum. Res. 30, 199–218.
Kahn, R.L., Wolfe, D.M., Quinn, R.P., Snoek, J.D., Rosenthal, R.A., 1964. Organizational Stress: Studies in Role Conflict and Ambiguity. John Wiley, Oxford, England.
Kruis, A., 2008. Management Control System Design and Effectiveness. PhD-thesis Nyenrode Business University.
Kruis, A.M., Speklé, R.F., Widener, S.K., 2016. The levers of control framework: an exploratory analysis of balance. Manag. Account. Res. 32, 27–44.
Landers, R.N., Behrend, T.S., 2015. An inconvenient truth: arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. Ind. Organ. Psychol. 8, 142–164.
Lee, F., Edmonson, A.C., Thomke, S., Worline, M., 2004. The mixed effects of inconsistency on experimentation in organizations. Organ. Sci. 15, 310–326.
Lewis, J.M., 2015. The politics and consequences of performance measurement. Policy Soc. 34 (1), 1–12.
Malmi, T., Brown, D.A., 2008. Management control systems as a package—Opportunities, challenges and research directions. Manag. Account. Res. 19 (4), 287–300.
Marginson, D., Ogden, D., 2005. Coping with ambiguity through the budget: the positive effects of budgetary targets on managers' budgeting behaviours. Account. Organ. Soc. 30, 435–456.
Marginson, D., McAuley, L., Roush, M., Van Zijl, T., 2014. Examining a positive psychological role for performance measures. Manag. Account. Res. 25, 63–75.
Masschelein, Stijn, Moers, Frank, 2020. Testing for complementarities between accounting practices. Account. Organ. Soc. 86.
Merchant, K.A., Van der Stede, W.A., 2017. Management Control Systems: Performance Measurement, Evaluation and Incentives, 4th edition. Pearson, New York.
Millsap, R.E., Zalkind, S.S., Xenos, T., 1990. Quick-reference tables to determine the significance of the difference between two correlation coefficients from two independent samples. Educ. Psychol. Meas. 50 (2), 297–307.
Moers, F., 2006. Performance measure properties and delegation. Account. Rev. 81, 897–924.
Nielsen, H., Kristensen, T.B., Grasso, L.P., 2018. The performance effects of complementary management control mechanisms. Int. J. Oper. Prod. Manage. 38 (11), 2124–2148.
OECD, 1994. Performance Management in Government: Performance Measurement and Results-oriented Management. PUMA Occasional Papers 3.
Otley, D., 2016. The contingency theory of management accounting and control: 1980–2014. Manag. Account. Res. 31, 45–62.
Otley, D.T., Berry, A.J., 1980. Control, organization and accounting. Account. Organ. Soc. 5, 231–244.
Ouchi, W.G., 1977. The relationship between organizational structure and organizational control. Adm. Sci. Q. 22, 95–113.

Ouchi, W.G., 1979. A conceptual framework for the design of organizational control mechanisms. Manage. Sci. 25, 833–848.
Petter, S., Straub, D., Rai, A., 2007. Specifying formative constructs in information systems research. Mis Q. 31, 623–656.
Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88, 879–903.
Pollitt, C., 2002. Clarifying convergence: striking similarities and durable differences in public management reform. Public Manag. Rev. 4, 471–492.
Rainey, H.G., Jung, C.S., 2015. A conceptual framework for analysis of goal ambiguity in public organizations. J. Public Adm. Res. Theory 25 (1), 71–99.
Rizzo, J.R., House, R.J., Lirtzman, S.I., 1970. Role conflict and ambiguity in complex organizations. Adm. Sci. Q. 15, 150–163.
Roberts, J., 2010. Designing incentives in organizations. J. Inst. Econ. 6 (1), 125–132.
Sharma, S., Mukherjee, S., Kumar, A., Dillon, W.R., 2005. A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. J. Bus. Res. 58, 935–943.
Simons, R., 1990. The role of management control systems in creating competitive advantage: new perspectives. Account. Organ. Soc. 15, 127–143.
Speklé, R.F., Verbeeten, F.H.M., 2014. The use of performance measurement systems in the public sector: effects on performance. Manag. Account. Res. 25, 131–146.
Speklé, R.F., Widener, S.K., 2018. Challenging issues in survey research: discussion and suggestions. J. Manag. Account. Res. 30 (2), 3–21.
Speklé, R.F., Widener, S.K., 2020. Insights on the use of surveys to study management control systems. Account. Organ. Soc. 86, 101184.
Speklé, R.F., Van Elten, H.J., Widener, S.K., 2017. Creativity and control: a paradox. Evidence from the Levers of Control Framework. Behav. Res. Account. 29 (2), 73–96.
Sprinkle, G.B., Williamson, M.G., Upton, D.R., 2008. The effort and risk-taking effects of budget-based contracts. Account. Organ. Soc. 33, 436–452.
Tanriverdi, H., Venkatraman, N., 2005. Knowledge relatedness and the performance of multibusiness firms. Strateg. Manage. J. 26, 97–119.
Tirole, J., 1994. The internal organization of government. Oxf. Econ. Pap. 46, 1–29.
Van de Walle, S., Sterck, M., Van Dooren, W., Bouckaert, G., Pommer, E., 2004. Public management. Public Sector Performance: an International Comparison of Education, Health Care, Law and Order and Public Administration. SCP Social and Cultural Planning Office, The Hague, pp. 234–270.
Van der Kolk, B., Van Veen-Dirks, P.M., ter Bogt, H.J., 2019. The impact of management control on employee motivation and performance in the public sector. Eur. Account. Rev. 28, 901–928.
Van Dooren, W., Bouckaert, G., Halligan, J., 2015. Performance Management in the Public Sector. Routledge.
Verbeeten, F.H.M., 2008. Performance management practices in public sector organizations: impact on performance. Account. Audit. Account. J. 21, 427–454.
Verbeeten, F.H.M., Speklé, R.F., 2015. Management controls, results-oriented culture, and public sector performance: empirical evidence on New Public Management. Organ. Stud. 36, 953–978.
Whithey, M., Daft, R.L., Cooper, W.H., 1983. Measures of Perrow's work unit technology: an empirical assessment and a new scale. Acad. Manag. J. 26, 45–63.