Management Accounting Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Management Accounting Research

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

Employee participation, performance metrics, and job performance: A survey study based on self-determination theory

Bianca A.C. Groen[a,*], Marc J.F. Wouters[a,b], Celeste P.M. Wilderom[c]

[a] University of Amsterdam Business School, P.O. Box 15953, 1001 NL Amsterdam, The Netherlands
[b] Karlsruhe Institute of Technology, Institut für Unternehmensführung, P.O. Box 6980, 76128 Karlsruhe, Germany
[c] University of Twente, School of Management and Governance, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 19 June 2015
Received in revised form 8 October 2016
Accepted 12 October 2016
Available online xxx

Keywords:
Employee participation
Self-determination theory
Operational performance metrics
Perceived measurement quality
Autonomous motivation
Job performance

ABSTRACT

Suitable and valid operational performance metrics are important means to translate an organization's strategy into action. However, developing high-quality operational metrics is challenging because such metrics need the right degree of context specificity to be meaningful to the managers and employees who will use them. We investigated whether managers consider metrics that have been co-developed with operational employees to be of higher quality and, in turn, whether they use these metrics more—and whether this use is linked to greater employee job performance. On the basis of self-determination theory, we investigated if different uses of performance metrics have different effects. We surveyed 86 pairs of operational employees and their immediate managers in various jobs and industries and tested our hypotheses with structural equation modeling. Results showed that when employees were involved in the development of performance metrics, managers perceived the metrics to be of better quality and employed those metrics more for evaluating and rewarding employees. Moreover, we found employees' performance was only higher when the metrics were used for evaluation purposes. We found no effect for using the metrics for monetary compensation or nonmonetary rewards. In sum, this study demonstrates that employee participation in the development of performance metrics has beneficial effects on the metrics' quality, and shows that the subsequent effect on job performance depends on how these metrics are used. We discuss implications for managers who want to ensure that the effect on employee job performance is positive when they involve employees in the development of operational performance metrics.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

A critical function of contemporary performance measurement systems is the translation of strategies into operational terms (Evans, 2004; Franco-Santos et al., 2007, 2012; Ho et al., 2014; Melnyk et al., 2004). Here, "operational" refers to activities performed by work-floor employees in line positions, such as operators and professionals. These individuals are directly involved in creating the actual products and/or services of the organization, whether they are operators in a factory, on-site construction workers, doctors working with patients, or university professors teaching classes.

Our focus on operational employees and their first-line supervising managers is a distinguishing feature of this study. The work of operational employees is often context-specific (McKinnon and Bruns, 1992), and the peculiarities of the work present challenges to designing performance metrics that are valid, suitable, and meaningful quantified indicators of strategically relevant work activities (Hopp et al., 2009; Jordan and Messner, 2012; Lillis, 2002). High-quality operational performance metrics need to consider key details of the idiosyncratic aspects of how work is carried out, such as links between activities, reasonable standards in a particular context, resources used for the activities, and ways particular events are recorded and generate data (Wouters and Roijmans, 2011).[1]

* Corresponding author.
E-mail addresses: b.a.c.groen@uva.nl (B.A.C. Groen), marc.wouters@kit.edu (M.J.F. Wouters), c.p.m.wilderom@utwente.nl (C.P.M. Wilderom).

[1] For example, in a case study on developing operational performance metrics, Groen et al. (2012) describe how one intricacy for designing a meaningful performance measure of electricity use in a production department was the exclusion of the battery charging station for the forklift trucks, because its electricity use could

http://dx.doi.org/10.1016/j.mar.2016.10.001
1044-5005/© 2016 Elsevier Ltd. All rights reserved.

Prior research has looked at various approaches companies take to deal with poor representational qualities of operational performance metrics. Some firms use metrics in combination with subjective performance evaluation (Gibbs et al., 2004), while others try to improve their measurement system, for instance by integrating it with other management systems or adding weights to the metrics (Lillis, 2002). Managers may adopt flexible ways of using indicators once they are in place (Jordan and Messner, 2012; Jørgensen and Messner, 2010), and combine these with complementary information (Hall, 2010). Research has shown that performance metrics become more valid and transparent to managers when these metrics include operational knowledge (Englund and Gerdin, 2015), and the current study assumes that operational employees know more operational details than their managers. Therefore, employee participation in the development of performance metrics can help in constructing more valid, suitable, and meaningful representations of operational performance. We investigate whether performance metrics that have been co-developed with employees are related to employee job performance, and whether this relationship is mediated by the quality of the metrics—as judged by the managers—and by managers' subsequent use of those metrics.

Employee participation—in the broadest sense—is a popular research theme in the field of management accounting (e.g., De Baerdemaker and Bruggeman, 2015; Derfuss, 2009; Jansen, 2015; Kruis and Widener, 2014; Webb et al., 2010). This paper focuses on employee participation in the development of performance metrics (*PM participation*), a topic that is increasingly examined in the management accounting literature. Research has shown participation may provide various benefits for individuals and organizations (Abernethy and Bouwens, 2005; De Haas and Algera, 2002; Groen et al., 2012; Hunton and Gibson, 1999; Kleingeld et al., 2004; Li and Tang, 2009; Wouters and Wilderom, 2008). Several studies have specifically focused on the performance effects of *PM participation* and found it can increase performance, sometimes indirectly (Abernethy and Bouwens, 2005; Groen et al., 2012, 2016; Hunton and Gibson, 1999; Kleingeld et al., 2004). However, all of these studies have investigated the relationship between *PM participation* and performance from the perspective of the participating individuals. Thus far, no study has provided insights as to what superiors think about such co-developed PMs and how they could use PMs to stimulate better employee performance.

This study contributes to the existing literature by investigating how employee job performance can be improved by involving operational employees in developing their own performance metrics. We focus on a different level in the organization: operational employees and their managers (i.e., first-line supervisors) rather than managers and the next higher level managers. Moreover, we focus on different roles: the reactions of the supervising managers rather than the reactions of participating individuals to involvement in the development of operational performance metrics. Investigation of the reactions of the supervising managers is important because eventually management makes decisions about giving employees the opportunity to participate in developing performance metrics, and about using the metrics to incentivize employees. The present research could help managers make better decisions in that regard. It investigates operational performance measurement through a behavioral lens and focuses on participation, different uses of the metrics, and employee job performance. Thereby it heeds the call to include behavioral theories in accounting research (Merchant et al., 2003), and builds upon recent insights that the impact of performance metrics depends upon how the metrics are used (e.g., Franco-Santos et al., 2012; Marginson et al., 2014; Van Veen-Dirks, 2010). This paper reports on whether involving employees in metrics development can increase managers' perceptions of the quality of performance metrics (*PM quality*), and how these metrics can be best used to increase employee job performance. We argue that managers should integrate the job-specific knowledge of operational employees into the performance metrics through employee participation. Such participation may lead to metrics that more aptly reflect the job performance of operational employees and, when used by managers in an appropriate way, may boost employees' job performance. We test three types of uses of these performance metrics as mediators in the relationship between *PM participation* and *employee job performance*.

Besides the overall contribution outlined above, the model tested in this study also aims to make two other contributions. First, the model focuses on *PM participation* as an antecedent of the perceived quality of the metrics. We acknowledge that managers could be concerned that participation may create difficulties (it may, for example, lead to disagreement and conflict, and may also consume much time) and that employees might use their knowledge to construct metrics that are more advantageous to themselves. We develop a nuanced argument as to why, nevertheless, both employees and managers may want to adopt *PM participation* to improve the validity of the operational performance metrics and to increase their sensitivity, precision, and verifiability. Therefore, in our model these measurement properties are not exogenously given, but are shaped in the development of the performance metrics. We test the theoretical model with survey data from pairs of managers and their subordinates from various operational types of jobs, organizations, and industries, and in line with the prediction, we find that *PM participation* of employees improves measurement properties as perceived by managers. Also, as expected, we find that managers make more use of the performance metrics to reward and evaluate employees when they perceive the metrics to be of better quality.

For the second additional contribution, we investigate consequences of perceived quality of performance metrics in more detail by examining three types of use of PMs by managers and, in turn, their differential effects on job performance. These types of use concern the importance of PMs for monetary compensation, nonmonetary rewards, and evaluation purposes. For this, we draw on the distinction between tangible rewards and verbal rewards that is central in research using self-determination theory (Cameron and Pierce, 1994; Deci et al., 1999a). Tangible rewards are concrete and explicit incentives that "are frequently offered to people as an inducement to engage in a behavior in which they might not otherwise engage" (Deci et al., 2001a, p. 4). We distinguish two different types of tangible rewards, and so our model includes the *use of PMs for monetary compensation* (such as salary increases or bonuses) and the *use of PMs for nonmonetary rewards* (such as increasing the chances of promotion). Using PMs in relation to verbal rewards means that PMs are brought up in expressions of appreciation and provide employees with information that helps them in their work. We include in our model the *use of PMs for evaluation purposes*, which refers to their use in performance evaluations, official performance ratings, and periodic discussions. The *use of PMs for evaluation purposes* focuses on substantive discussion about what has been driving performance and which actions could be taken, whereby PMs inform this process: "In the periodic evaluation, the

not be influenced by the employees. As another example, a performance measure for on-time grading of examinations in a university context could be complicated by many operational peculiarities that would have to be incorporated in the metrics, such as: What does on-time mean and how can it be measured if, for a given course, an individual student's final grade can only be determined and reported after that student has handed in a final assignment, but the deadline for that assignment has been set several weeks after students have taken the written examination for that course? What counts as the compliance date—the date on which grades are unofficially published on-line by the lecturer? Or is it the date on which the lecturer sends the grades to Dean's the office, or the date on which that office is able to enter those grades into the official system?

Please cite this article in press as: Groen, B.A.C., et al., Employee participation, performance metrics, and job performance: A survey study based on self-determination theory. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.10.001

production manager and the superior discuss the functioning of the production department and the situation in which the department operates with the help of a set of performance measures" (Van Veen-Dirks, 2010, p. 142).

We draw on self-determination theory to develop differential hypotheses about the performance consequences of the three uses of PMs. A key theme of self-determination theory is that the use of performance metrics for incentivizing employees is not without problems, because providing incentives may have no or even a negative effect on performance (Ashton, 1990; Bonner et al., 2000; Bonner and Sprinkle, 2002; Fessler, 2003; Kunz and Linder, 2012; Kunz and Pfaff, 2002; Speklé and Verbeeten, 2013; Wong-On-Wing et al., 2010). Our study is unique in investigating the potentially offsetting performance effects of using metrics toward three different purposes: for monetary compensation, for nonmonetary rewards, and for evaluation. Prior studies on the relationship between incentives and performance typically studied very specific and often experimental situations. In contrast, this study focuses on a broad sample of employees. In line with the hypotheses, we find that *PM participation* by operational employees has positive job performance effects if the resulting high-quality performance metrics are used for evaluation purposes. We find no relationship with employee job performance if the performance metrics are used for monetary compensation or nonmonetary rewards, which challenges the almost universal belief in the positive effects of creating incentives based on using performance metrics.

2. Theory and hypotheses

This study investigates how involving operational employees in developing performance metrics can lead to better employee job performance. It specifically looks at how *PM participation* improves managers' perceived quality of operational performance metrics and how, in turn, managers use such metrics in various ways with the aim of elevating *employee job performance*. Fig. 1 summarizes the study's model and includes definitions of the constructs. This section introduces the model and explains how it draws on self-determination theory.

2.1. PM participation, PM quality, and use of PMs

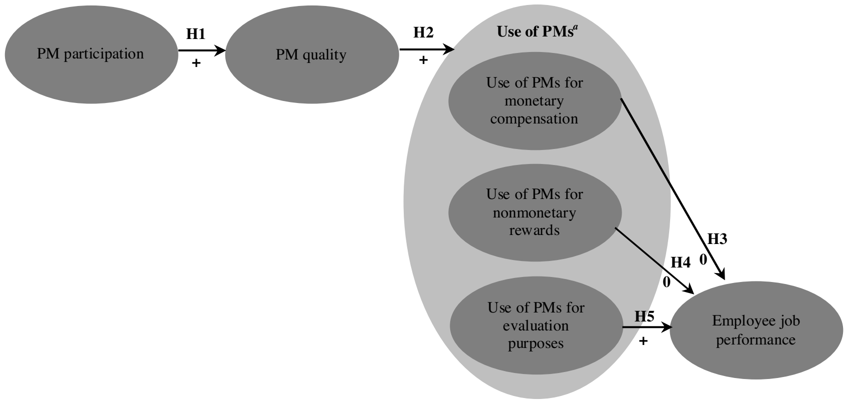
We investigate managers' perceptions of the quality of operational performance metrics (*PM quality*) after employees have participated in developing them. Performance metrics are quantitative expressions of how the operational employees have conducted their tasks. Examples are client satisfaction, process efficiency, timeliness, and quality of work completed. *PM quality* concerns measurement properties of the operational performance metrics and we define it as the extent to which managers find the performance metrics to be sensitive to the actions of their employees, precise in measuring relevant aspects of their employees' performance, and verifiable (see Fig. 1 for definitions of all the constructs). We define the construct "employee participation in developing performance metrics," or *PM participation*, as the extent of influence employees feel they have had on the design of the performance metrics they are measured by. *PM participation* encompasses the employees' influence on the development of *all aspects* of performance metrics during all the phases: design, implementation, and maintenance of the metrics (Bourne et al., 2000; Kennerley and Neely, 2003; Neely et al., 1997, 2002). *PM participation* goes far beyond the setting of targets: employees may influence many design factors, including the conceptualization of the metrics, identifying required data, adapting IT systems, designing graphs and tables, and even producing the periodic performance reports (Groen et al., 2012). *PM participation* is distinct from the interactive use of existing performance measurement systems, which has been regularly investigated in the literature (e.g., Bisbe and Otley, 2004; Henri, 2006; Koufteros et al., 2014; Widener, 2007). The interactive use of performance metrics refers to how managers and employees use existing PMs in their communication (Ferreira and Otley, 2009), whereas *PM participation* is about how managers and employees work together to design and implement a new or modified PMs.

PM participation involves information sharing between subordinates and superiors (Shields and Shields, 1998). Employees are assumed to have better knowledge about operational practices and processes than their managers (Kim et al., 2014), and during the PM participation process, employees may communicate some of this job-specific information and incorporate it into the metrics (Chow et al., 1988; Merchant, 1981; Nouri and Parker, 1998; Shields and Shields, 1998; Wouters and Roijmans, 2011). However, as not all *PM participation* initiatives are successful (Kruis and Widener, 2014), some managers may be reluctant to try to improve the performance metrics by involving employees in the metrics' development. They could be concerned that participation could be time-consuming and that it may lead to disagreement and conflict (Poon et al., 2001). Alternatively, in line with the traditional economics-based view, they may be afraid employees would use such an opportunity to influence the development of new performance metrics in a direction that only benefits them. Managers might fear that the performance metrics will be useful and insightful only for the employees, but make the performance of employees less transparent for managers, or even favorably biased. Examples of such reasoning can be found in the budgeting literature, in which budgetary participation is traditionally seen as an important antecedent of budgetary slack. De Baerdemaker and Bruggeman (2015) provide a recent overview this literature.

Several counterarguments suggest that managers do not have to be overly concerned about employee participation. First, the degree to which operational employees may affect the shape and content of the performance metrics is limited. Maintaining congruence between the performance metrics and the organizational strategy remains the task of management, and managers can preserve this congruence by setting boundaries as part of the PM participation process. For example, they might define the strategic priorities the performance metrics should reflect (Groen et al., 2012) or take part in the development process to ensure the co-developed performance metrics align with the organization's strategy. Moreover, managers also have private information, and employees do not know exactly what managers know and do not know. This lack of knowledge limits how much information employees can keep private and use to their advantage (cf., the economic models of Baiman and Evans, 1983; Christensen, 1982; Penno, 1984).

More fundamentally, we propose that employees have limited interests in abusing *PM participation*, since good measurement properties may also be beneficial for the employees themselves (Jordan and Messner, 2012).[2] Employees need sensitive performance metrics because they want their work efforts to be recognized and therefore reflected in these metrics, especially when they are performing well. Hence, we argue that employees benefit from sensitive metrics in that they may prefer performance metrics that recognize good performance and that do not disclose bad performance. A similar argument can be made for the precision of the performance metrics. Employees (who are typically assumed to be risk-averse) do not want noisy metrics that occasionally attribute windfalls to them but at other times blame them

[2] Jordan and Messner (2012) investigated managers' reactions to the quality of performance metrics used by top management to assess the managers' performance. Nonmanagerial employees are likely to have similar reactions to the ways in which their performance is being measured.



Construct short name	Construct definition
PM participation	The extent of influence employees feel they have had on the design of the performance metrics they are measured by
PM quality	The extent to which managers find the performance metrics to be sensitive to the actions of their employees, precise in measuring relevant aspects of their employees' performance, and verifiable
Use of PMs[a]	The extent to which managers find it important to use the metrics for monetary compensation, nonmonetary rewards, and evaluation purposes
Use of PMs for monetary compensation	The extenstion to which managers find the performance metrics to be important for giving their employees salary increases, bonuses, or extras
Use of PMs for nonmonetary rewards	The extent to which managers find the performance metrics to be important for increasing their employees' chance of promotion or authority within the organization
Use of PMs for evaluation purposes	The extent to which managers find the performance metrics to be important in performance evaluations, official performance ratings, and periodic discussions
Employee job performance	The extent to which employees meet their job requirements according to their manager

[a] Use of PMs is a second-order construct

Fig. 1. Hypothetical model and construct definitions.

for bad performance they could not prevent. Employees also find verifiability or transparency to be a desirable property, as they will be unlikely to want their performance to be assessed with metrics containing questionable, untrustworthy data that cannot be checked (Englund and Gerdin, 2014; Jordan and Messner, 2012).

Furthermore, employees are generally more honest in participation situations than would be expected on the basis of agency theory (Brown et al., 2009; De Baerdemaker and Bruggeman, 2015). Also, employees' preferences for fairness may reduce the risk of getting inferior metrics out of a participation process. First, the participation process in itself increases employees' perceived fairness (Derfuss, 2009; Lau and Tan, 2005, 2006; Thibaut and Walker, 1975; Wentzel, 2002), making employees less likely to game the system because people tend to respond in a reciprocal way (Fehr and Falk, 2002; Fehr and Gächter, 2000; Langevin and Mendoza, 2013; Williamson, 2008). Second, since employees perceive high-quality performance metrics to be fairer (e.g., Burney et al., 2009; Langevin and Mendoza, 2013), their preferences for having a fair evaluation system in place will likely lead them to develop better quality performance metrics—also in the eyes of their managers. In other words, we argue that employees prefer to be involved in the development of performance metrics to improve the measurement properties rather than risk the imposition of metrics that poorly reflect the operational characteristics of their work. If employees do not participate, managers could use performance metrics that have inferior measurement properties, with low sensitivity to the employees' improvement efforts, high noise, and a basis in inappropriate data.

The preceding reasoning suggests that both employees and managers may benefit from better performance metrics, and the quality of performance metrics is therefore likely to increase when they develop the metrics together. If the resulting performance metrics are indeed of higher quality, the information from these metrics will be more valuable, and managers are more likely to use the metrics for various purposes, such as rewarding and evaluating employees (Banker and Datar, 1989; Feltham and Xie, 1994; Gibbs et al., 2009; Holmström, 1979; Van Veen-Dirks, 2010). We define *use of PMs* as the extent to which managers find it important to use the metrics for monetary compensation, nonmonetary rewards, and evaluation purposes. *Use of PMs* is a second-order construct that is reflected by three first-order constructs (see Fig. 1): *use of PMs for monetary compensation* is the extent to which managers find the performance metrics to be important for giving their employees salary increases, bonuses, or extras; *use of PMs for nonmonetary rewards* is the extent to which managers find the performance metrics to be important for increasing their employees' chances of promotion or authority within the organization; and *use of PMs for evaluation purposes* is the extent to which managers find the performance metrics important in performance evaluations, official performance ratings, and periodic discussions.

In sum we expect:

H1. Managers are more positive about the performance metrics' quality when employees have had more influence in their design.

H2. Managers make more use of the performance metrics when they find the performance metrics to be of good quality.

2.2. Self-determination theory and effects of distinct uses of PMs

The idea that incentives should have a positive effect on employee job performance has long dominated the literature (Jenkins et al., 1998; Stajkovic and Luthans, 1997, 2003). Agency theory presumes that employees' self-interest differs somewhat from the interest of the organization, and that control problems arise because actions are not observable and employees have private information. Therefore, managers should use performance metrics—imperfect surrogates of behavior—and incentives to make sure their employees work in a way that contributes to the overall organizational objectives (Jensen and Meckling, 1992). In line with expectancy theory, agency theory also suggests that people want to perform better if they expect to receive a reward for good performance, because better performance will then benefit them personally (Bonner and Sprinkle, 2002). Moreover, incentives may serve as screening mechanisms that can help attract more able employees, because individuals with higher skill levels are more likely to choose compensation contracts with higher performance incentives. Sprinkle's review includes a section of this stream of research (Sprinkle, 2003, pp. 292–293). Many studies have found support for the assumption that the use of incentives has a positive relationship with employee performance (e.g., Banker et al., 2000; Lazear, 2000; Presslee et al., 2013; Shearer, 2004). We do not discuss these arguments in detail as the contribution of our study lies elsewhere.

At the same time, many studies find no or a negative relationship between the use of incentives and employee performance (e.g., Bonner et al., 2000; Bonner and Sprinkle, 2002; Fessler, 2003; Holmås et al., 2010; Speklé and Verbeeten, 2014). We adopt a recent view of human motivation offered by self-determination theory, which is becoming increasingly popular also in the management accounting literature (Christ et al., 2012; De Baerdemaeker and Bruggeman, 2015; Kunz, 2015; Kunz and Linder, 2012; Kunz and Pfaff, 2002; Stone et al., 2010; Wong-On-Wing et al., 2010). The theory recognizes that employees' self-interest does not necessarily differ from the interest of the organization, because individuals seek to fulfill also their personal (often nonmonetary) needs, such as autonomy or recognition for their competence, and thus focus more on intrinsic rewards (Hall, 1996, 2004; Segers et al., 2008).

Self-determination theory sees employee motivation as a continuum from autonomous motivation to controlled motivation (Ryan and Deci, 2000). Under autonomous motivation, people feel they have the choice to do something and behave in ways to satisfy their personal needs, whereas under controlled motivation they feel they are being pressured to act to satisfy an external demand (Gagné and Deci, 2005; Ryan and Deci, 2000). Employees who have interesting tasks will already be autonomously motivated to do the tasks that are expected of them, which would preclude the need for incentives to stimulate them to perform well (Prendergast, 2008). Worse, giving extrinsic rewards may reduce autonomous motivation: when managers reward the completion of a certain task, they signal that the employees' current involvement and competence are inadequate or that the task should be seen as unattractive, because otherwise no extrinsic reward would be necessary (Bénabou and Tirole, 2003; Frey and Jegen, 2001). Most probably, employees' motivation will shift along the continuum from the autonomous motivation for the required task to controlled motivation (Bénabou and Tirole, 2003; Deci et al., 1999a; Kunz and Linder, 2012). This "crowding out" of autonomous motivation is dangerous, because in the long run controlled motivation is less powerful than autonomous motivation: once controlled motivation has gone up at the expense of autonomous motivation, this shift can hardly be undone (Bénabou and Tirole, 2003; Gneezy et al., 2011). Thus, the use of incentives can decrease performance (Bénabou and Tirole, 2003; Deci et al., 1999a; Falk and Kosfeld, 2006; Frey, 1994; Frey and Jegen, 2001; James, 2005; Lourenço, 2016; Prendergast, 2008; Wong-On-Wing et al., 2010).

To understand how various ways of using performance metrics may have differing effects on autonomous motivation and eventually employee performance, we must dive a bit deeper into self-determination theory.[3] The theory views autonomous motivation as a result of three innate psychological needs: the need for autonomy, relatedness, and competence (Deci and Ryan, 2000; Ryan and Deci, 2000). The need for autonomy is relevant to the *use of PMs for monetary compensation* and the *use of PMs for nonmonetary rewards*. People feel less autonomous when they perceive incentives as oppressing and controlling—as is typically the case with monetary or nonmonetary rewards that are contingent on quantitatively measured performance (Deci et al., 1999a; Fehr and Falk, 2002; Frey and Jegen, 2001). An example is cash bonuses based on meet or beat targets, or promotion only if one is ranked #1 among all sales managers on the basis of quarterly sales growth. Hence, *use of PMs for monetary compensation* as well as *use of PMs for nonmonetary rewards* may impede the satisfaction of the need for autonomy (Drake et al., 2007), thus decreasing the existing autonomous motivation of employees (Christ et al., 2012) and lowering their job performance. We distinguish monetary compensation from nonmonetary rewards—although both are expected to have basically the same effects—because nonmonetary rewards are often less concrete and less explicitly anticipated, which implies their harmful effect on autonomous motivation may be smaller (Gneezy et al., 2011). Meta-analyses have found mostly an overall negative effect of tangible rewards on autonomous motivation for interesting tasks (Deci et al., 1999a, 2001a; Cameron et al., 2001; Cameron and Pierce, 1994; Eisenberger and Cameron, 1996).

The need for competence is also relevant to the use of *PMs for monetary compensation* and the use of *PMs for nonmonetary rewards.* Using performance measures in these ways provides information to employees about how well they do relative to others and relative to targets, and receiving such information may help satisfy the need for competence and hence increase autonomous motivation (Cameron et al., 2005; Deci et al., 1999a). Cameron et al.'s (2001) meta-analysis has found some evidence for such a positive relation between tangible rewards and autonomous motivation, for example, for rewards offered for exceeding others (i.e., relative performance evaluation) and for uninteresting tasks.[4]

The needs for relatedness and for competence are relevant in relation to *use of PMs for evaluation purposes*. The need for relatedness is satisfied through the social interaction that takes place

[3] From the beginning, there has been much debate around self-determination theory, especially between the research groups of Deci, on the one hand, and Cameron, on the other hand. Both groups include mostly the same studies in their samples for their *meta*-analyses and, thus, most results are basically the same, at least at the level of verbal versus tangible rewards, which is relevant for our study. Compare, for example, Fig. 2 in Eisenberger and Cameron (1996) with Figs. 1 and 2 in Deci et al. (1999a). However, Deci's group emphasizes the detrimental effect of extrinsic rewards on autonomous motivation and eventually performance (Deci et al., 1999a, 1999b, 2001a, 2001b), whereas Cameron's group mainly emphasizes the situations in which extrinsic rewards have positive effects on (either autonomous or controlled) motivation and hence performance (Cameron, 2001; Cameron et al., 2001; Cameron and Pierce, 1994; Eisenberger and Cameron, 1996). Many of these seeming differences originate from focusing on different measurements of autonomous motivation. For example, the research group of Deci accentuates those results that are based on measurement of autonomous motivation using "free choice" which means observing research participants' time spent spontaneously on the experimental task after the experiment has "officially" ended. Yet, Cameron's group highlights results that are based on measurement of autonomous motivation using self-reported task interest. This focus on different measurements appears to be the main source of divergence in the overall conclusions that both groups underscore.

[4] Eisenberger et al. (1999) argue that also the need for autonomy may be fulfilled by tangible rewards. Their arguments rely on the notion of "locus of control," which they see as a manifestation of autonomy. However, we agree with Deci et al. (1999b) that autonomy does not regard "locus of control" but "locus of causality." Locus of causality is "the degree to which people experience their behavior to be volitional and freely chosen rather than coerced or seduced by desired outcomes" (Deci et al., 1999b; p. 693), which is similar to autonomy. Conversely, locus of control concerns the informational aspect of rewards, and hence should be seen as a prerequisite for the fulfillment of the need for competence rather than for autonomy.

during the evaluation (Kunz and Linder, 2012). The satisfaction of the need for relatedness is likely to increase the employee's existing autonomous motivation, and, in turn, raise the employee's job performance. The need for competence is satisfied during the substantive discussion and evaluation of performance, because in such discussions performance metrics are seen as informational (Deci et al., 1999a; Fehr and Falk, 2002; Frey and Jegen, 2001). In turn, satisfaction of the need for competence is likely to boost employees' autonomous motivation, and consequently may raise employees' job performance. Correspondingly, meta-analyses show that autonomous motivation increases when verbal rewards are used (Deci et al., 1999a; Eisenberger and Cameron, 1996). By verbal rewards they mean mainly positive feedback (Deci et al., 1999a), which has similarities with the operationalization of our PMs for evaluation purposes, although the evaluations themselves may not necessarily be positive or favorable. Meta-analyses have found an overall positive effect of verbal rewards on autonomous motivation (Deci et al., 1999a, 2001a; Cameron et al., 2001; Cameron and Pierce, 1994; Eisenberger and Cameron, 1996).

The studies included in meta-analyses were all experiments, mainly in educational settings. It is unclear whether we should expect similar findings in the average daily work life as is under investigation in the current study, in particular because in daily life there is likely a high diversity of how interesting tasks are and, thus, how much initial autonomous motivation will be present. Rewards may be positive for non-interesting tasks (Cameron et al., 2001), which is consistent with the classical agency theory argument. All in all, it seems safe to expect that in certain circumstances, extrinsic rewards may have positive effects on employee job performance and in other circumstances the effect may be negative. Hence, we formulate null hypotheses for the effects of *use of PMs for monetary compensation* and *use of PMs for nonmonetary rewards* on *employee job performance* (H3 and H4). The literature is less divided about the expected effects of using performance metrics for evaluation purposes. Such use is expected to maintain or increase autonomous motivation and consequently employee job performance (H5).

H3$_0$. There is no effect on job performance of employees when managers use the performance metrics for giving monetary compensation.

H4$_0$. There is no effect on job performance of employees when managers use the performance metrics for giving nonmonetary rewards.

H5. The job performance of employees is higher when managers use the performance metrics for evaluation purposes.

Overall, H1–H5 point toward the expectation that *PM participation* leads to better employee job performance if the performance metrics are used for evaluation purposes, whereas there is no clear expectation with regard to job performance effects of their use for monetary compensation or nonmonetary rewards. Therefore, we also hypothesize:

H6$_0$. No indirect relationship exists between *PM participation* and *employee job performance* via *PM quality* and *use of PMs for monetary compensation.*

H7$_0$. No indirect relationship exists between *PM participation* and *employee job performance* via *PM quality* and *use of PMs nonmonetary rewards.*

H8. A positive indirect relationship exists between *PM participation* and *employee job performance* via *PM quality* and *use of PMs nonmonetary rewards.*

3. Method

Because publicly available archival data for testing the hypotheses are not available, we collected our own data, using a survey method and following the framework of Van der Stede et al. (2005) for assessing the quality of survey research. This framework divides the requirements for survey research into five categories. First, researchers should have a specific research objective in mind to guide their study. As we reported in Section 1, our objective here is to discern whether metrics that have been co-developed with operational employees are considered to be of higher quality and, in turn, are used more by managers and whether this use is linked to higher employee job performance. Second, researchers should define their population and be clear about their sample to be able to know what inferences can be drawn from the study (see Section 3.1). Third, researchers should consider the survey questions and other research method issues necessary to judge the internal validity of the study (see Section 3.2). Fourth, the collected data should be accurate. Thus we describe the followed practical procedures when gathering the data (in Section 3.1) and report how we checked for possible problems in the dataset (in Section 3.3). And fifth, researchers should report how they ensured these requirements were met. The section below presents in detail the methods used.

3.1. Data collection and respondents

We tested our hypotheses with a survey among pairs of operational employees and their supervisory managers in the Netherlands. Respondent pairs had to meet three criteria: (1) the employees had to carry out operational activities of their organization; (2) the employees must have worked in their current functions for at least one year; and (3) the organization must have performance metrics for employees' performance. As these criteria make distinguishing the target population from the general population difficult if not impossible, we were precluded from using a random sampling technique to develop a sampling frame. A solution to this obstacle is to use respondent-driven (or "snowball") sampling, which involves asking each potential respondent for contact details of other people who meet the criteria. When the population is "hidden," respondent-driven sampling leads to samples that are just as good, and often even better, than those resulting from random sampling methods (Salganik and Heckathorn, 2004). Numerous studies have used this sampling method for that reason (e.g., Dalton et al., 2014; Lander et al., 2013; Neu et al., 2014; Raschke et al., 2014). Although some scholars suspected the sample from such a method would be dependent upon the starting point of the search, Salganik and Heckathorn (2004) substantiated that a respondent-driven sample is asymptotically unbiased, independent of one's starting point. Furthermore, Derfuss' (2009) meta-analytic study confirmed that the results of studies with random versus non-random samples are comparable.

We began our search for respondents by contacting people from our own network, including an alumni association and a consulting firm that sent our request to all of their contacts, mainly by email. Additionally, we collaborated with two professional associations by organizing two seminars on developing useful performance metrics. All seminar participants had completed a survey in advance to make the seminar more relevant for their own situations. Respondents were further recruited after we published papers in three Dutch professional journals describing how to develop performance metrics jointly with operational employees. In these articles we recommended that readers complete a survey if they wanted to know whether this approach was relevant to them. We promised participants a free copy of the research report with their personal scores and the averages of other organizations for comparison with their scores. Finally, we promised all respondents an invitation to a seminar during which they could exchange their experiences with performance measurement, and in which they would receive recommendations based on the results of earlier research about the co-development of operational performance metrics.

Please cite this article in press as: Groen, B.A.C., et al., Employee participation, performance metrics, and job performance: A survey study based on self-determination theory. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.10.001

Table 1
Respondent characteristics.

Characteristic	Employees	Managers
Gender	73% male 27% female	88% male 12% female
Education	38% low-level high school 35% high-level high school/BSc 27% MSc or higher	21% low-level high school 49% high-level high school/BSc 30% MSc or higher
Age	Mean = 39 (SD = 9.7)	Mean = 45 (SD = 6.8)
Departmental tenure	Mean = 6.2 (SD = 6.2)	Mean = 7.8 (SD = 6.2)
Span of control	N/A	Mean = 35 (SD = 52)
Size of the organization (# staff)	Mean = 5706 (SD = 23,748) min 7, max 150,000	Same as for employees, as manager- employee pairs belong to the same organization
Job type of the employee	30% operator 43% service employee 27% professional	

To find suitable pairs of respondents comprising an employee and his or her manager, we used the following procedure. Employees who agreed to participate completed the online survey and provided us with their manager's contact details. Managers who agreed to respond to the survey provided us with the contact details of employees who met the pre-specified criteria. After one employee (randomly chosen by us) completed the survey, we contacted the manager again to complete an online survey. The name of the employee who completed the employee survey was provided to the manager, to make sure the manager would explicitly fill in the questionnaire for the performance of that particular employee. We made clear to all survey respondents that their answers would be kept strictly confidential. Hence, the employees and managers received only a feedback report of their personal results and no report of the results of the other half of their pair.

The survey was completed by 86 pairs of respondents. At first, 21 employees and 74 managers indicated they met the inclusion criteria and wanted to complete the rest of the survey, potentially giving us 95 pairs. All 21 employees completed the survey but only 15 of their managers. For 74 managers who wanted to take part in the survey, all of their randomly selected subordinate employees completed the survey, and subsequently 71 of the 74 managers completed their particular part. Thus, in total, we received data from 15 + 71 = 86 complete pairs (89%). Table 1 gives an overview of the respondents' characteristics.

3.2. Survey instrument

We pretested a pilot version of the survey among 17 employees who did not participate in the final survey but had characteristics similar to those of the employees in the population (Anderson and Gerbing, 1991). Three pretest methods were used simultaneously. First, Anderson and Gerbing's (1991) item-sort task was executed to determine how well the items measured the constructs. Second, Hak et al.'s (2008) three-step test-interview method was used to find out more about the respondents' actual response behavior. Participants were asked to complete the survey and, at the same time, say aloud what they were thinking. Third, Cronbach's alphas of the scales were calculated and principal component analyses were performed. These three pretest methods provided triangulated data that helped shorten the survey and suggested the metrics were valid.

The final survey items measure the constructs shown in Appendix A. Respondents rated all items on a seven-point fully anchored Likert scale: (1) totally disagree, (2) disagree, (3) moderately disagree, (4) neutral, (5) moderately agree, (6) agree, (7) totally agree. All the items were in Dutch. Employees completed the survey questions regarding *PM participation*, whereas managers completed the questions regarding *PM quality*, *use of PMs*, and *employee job performance*. The use of respondent pairs increased the construct validity of our variables, because each variable was measured at the relevant person. Furthermore, the design reduced potential common source bias. Additionally, we followed conventional guidelines to reduce common method bias (Podsakoff et al., 2003). First, each survey page started with a brief introduction and contained only the items belonging to one construct—an approach that improves the quality of the data because it helps respondents better understand the items (Frantom et al., 2002). The constructs were also measured in an order that differed from that of the model. Confidentiality of the answers was emphasized, and the scale items had been carefully constructed and pretested. Moreover, we controlled for common method bias statistically by adding a latent common method variable to the model (Podsakoff et al., 2003).

PM participation is measured with five items (shown in Appendix A) that tapped the degree of influence employees feel they had on the design of the performance metrics in use to measure their performance (Abernethy and Bouwens, 2005). Cronbach's alpha is 0.94. We allowed for covariance of the error terms of two items ("I have/had influence on ongoing modifications to the design of the performance metrics" and "I have/had influence on the maintenance of the performance metrics") because, especially in the Dutch language, these two items are very similar in that they both deal with adjusting the performance metrics when they are already in use.

PM quality is assessed with items that measure the extent to which managers find the performance metrics sensitive to the actions of their employees, how precise the performance metrics are in measuring relevant aspects of their employees' performance, and whether they are verifiable (Moers, 2006). The pretest helped us add and delete items from the original scale to make the scale relevant for operational employees. More items were included in the pretest than were intended for the final analyses because which items of the original scale were valid was unclear. The properties of these items were analyzed before testing the hypotheses to ensure no false positives were created by "cherry-picking" items. A varimax rotated principal component analysis of the final data yielded three factors. The core of our PM quality operationalization consists of one of these factors, which with a mixture of precision, sensitivity, and verifiability items best reflects the definition of *PM quality*.[5] These items were all part of the original scale. Cronbach's alpha is 0.72. Several robustness checks were carried out to see whether the results would have been similar if the deleted items were added to the analyses. The path coefficients and their significance levels did not change when the other items or complete factors were added, but the model fit was reduced.

Use of PMs is a second-order construct that is reflected by three first-order constructs: *use of PMs for monetary compensation*, *nonmonetary rewards*, and *evaluation purposes*.[6] This second-order structure is a natural representation of the use of PMs since the

[5] One of the other factors contained the four negatively formulated (recoded) precision items from the original scale. The fact they loaded on a separate factor is surprising since they should measure the same as the two other precision items. Apparently their format rather than their contents makes them load on a different factor. Therefore, these items were deleted. The other deleted factor consisted of three items that capture making more of an effort in the employee's job leads to better performance on the performance metrics. These items deal with employee input, while performance metrics should deal with employee output.

[6] A graphical representation of the structural model, including this second-order construct and its underlying first-order constructs, is given in Appendix B.

three types of use tend to go hand in hand (as, for example, when someone gets a raise because of a promotion, or when the same metrics are used for both evaluation purposes and determination of rewards (Grafton et al., 2010)). Also, incorporating this second-order structure avoids multicollinearity problems (Koufteros et al., 2009). Any commonality in variance is captured in the second-order construct use of PMs, and the three specific types of use only reflect their unique variance.[7] These three first-order constructs are measured by three subscales on the use of performance metrics and are all collected from the managers (Moers, 2006). The first two are measured with two items and the third with three items. Originally, the *use of PMs for evaluation purposes* scale consisted of four items. We excluded the fourth item ("I attach very high importance to the performance indicators in periodic performance reports") because its factor loading was below 0.40, and the pretest showed that respondents perceived this item as a measure for *PM quality*. The Cronbach's alpha is 0.84 for *use of PMs for monetary compensation*, 0.88 for *use of PMs for nonmonetary rewards*, and 0.89 for *use of PMs for evaluation purposes*.

Employee job performance was measured with the well-known scale for in-role job performance originally developed by Williams (see Williams and Anderson, 1991) and later revised and abbreviated by Podsakoff and MacKenzie (1989). It measures the extent to which managers perceive employees as meeting their job requirements. It considers job performance in general, from the perspective of the manager. This scale is relevant in this study because increasing the extent to which employees meet their job requirements is precisely the behavior that performance metrics usually would need to stimulate (Williams and Anderson, 1991). Further, the scale correlates highly with objective measures of performance (Burney et al., 2009). Using objective measures of performance *per se* would have been problematic, because the large diversity in jobs would make the measures incomparable. The broad applicability of the scale allows comparison of employee performance in various jobs and industries. Cronbach's alpha is 0.91.

The control variables used in this study are gender, educational level, age, departmental tenure and job type of the employee, size of the organization and the department, length of time the performance metrics have been implemented, information asymmetry, and delegation. Demographic variables are included because previous studies (e.g., Ali and Davies, 2003) have shown they may explain differences in *employee job performance* and conceivably explain variance in, for instance, *PM participation*, which makes them potentially confounding. Length of time the performance metrics had been in use was included as a potential confounding variable because it may be related to *PM participation* (if the PMs have been in place for a longer time employee influence is less likely), *use of PMs* (PMs in place for a longer time are more likely to be institutionalized and thus used by the managers), and perhaps also *employee job performance* (if the PMs have been there for a longer time, employees may have gotten used to the measures and will therefore better know how to perform well in their jobs). Employees self-reported their gender, age, educational level, departmental tenure, length of time the performance metrics had been implemented, information asymmetry, and delegation. The manager provided the size of both the organization and the department. A job-type dummy was coded by the researchers: the names of the organizations involved in the study were mentioned in respondents' email addresses, which afforded the possibility to determine what industry they were active in. The organizations were coded as "manufacturing," "mass service," or "professional service." Because this study focused on operational employees only, these codes also indicate whether the employee was an operator, service employee, or professional.

Information asymmetry is the amount of employees' specific knowledge compared to that of their manager. This variable is included because the involvement of employees in the development of better performance metrics is only relevant if the employees have particular knowledge that complements what their managers know. It is measured with Dunk's (1993) scale, which employees completed. Cronbach's alpha is 0.93.

Delegation (of decision rights) is included because—together with *PM quality* and *use of PMs*—it is traditionally seen as one of the three primary components of organizational design that are supposed to be mutually complementary (Ortega, 2009; Widener et al., 2008). Furthermore, the organizational behavior literature proposes that delegation is related to *employee job performance* since it is a form of autonomy that increases autonomous motivation and therefore eventually employee job performance (Deci and Ryan, 2000). *Delegation* is defined as giving employees the freedom and right to decide and act on their own (Mills and Ungson, 2003). Employees completed the scale of Thomas and Tymon (1993). We deleted one item ("I make my own choices without being told by management") because this item severely reduced the model fit and was already problematic in the pretest.[8]

3.3. Statistical analyses

The data were analyzed with structural equation modeling, using maximum likelihood estimation with AMOS 18. Structural equation modeling was chosen over a series of regression analyses because it allows for simultaneous analyses of multiple and interrelated relationships of latent variables. That is, all relationships in the model, as well as the measurement model, are estimated at once. We used covariance-based structural equation modeling because the goal is to test formal hypotheses on the basis of a theory (Anderson and Gerbing, 1988; Peng and Lai, 2012), and a recent comprehensive series of Monte Carlo simulations showed that, compared to variance-based, covariance-based structural equation modeling is more accurate for smaller samples, thanks to its compensation for measurement error (Goodhue et al., 2012).[9] Most assumptions of covariance-based structural equation modeling were met: the data showed no collinearity and no outliers ($p < 0.001$), and were univariately normal. One assumption that was not completely met was that of multivariate normality. As we found multivariate kurtosis, we used bootstrapping to ensure that this did not influence the results (Kline, 2011, p. 177).[10] We did not experience any estimation problems.

We followed the two-step approach of Anderson and Gerbing (1988), which makes determining the source of poor fit relatively straightforward. Moreover, it enables the detection of interpretational confounding (Burt, 1976), which means the empirical definitions of the constructs (i.e., factor loadings) change depending on the structural model (Kline, 2011). The first step consisted of a confirmatory factor analysis to estimate the fit of the measurement model. That is, all items in the measurement model were

[7] Using a second-order model with structural equation modeling is similar to using residuals obtained from regressing each use of PMs on the other two (cf. Hansen and Van der Stede, 2004; p. 429)) and has the additional advantage that the whole model is estimated at once.

[8] In hindsight, we think the Dutch version of this item could easily be misread as: "although the management gives me no freedom in my work, I will create this freedom myself." What we meant to assess was that management gives employees the freedom to make autonomous choices, that is, without consulting their manager.

[9] Our sample consisted of 86 respondent pairs, whereas with four latent constructs and 21 indicators, a sample size of 116 or more is recommended (Westland, 2010).

[10] Whenever we used bootstrapping, we used 1999 bootstrap samples and bias-corrected confidence intervals set to 95%.

Please cite this article in press as: Groen, B.A.C., et al., Employee participation, performance metrics, and job performance: A survey study based on self-determination theory. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.10.001

only allowed to load on their intended factor, and the factors were allowed to freely correlate with each other.

Once the measurement model was adequate, the structural model was analyzed. We began with a model containing only the hypothesized relationships, hence without control variables. This exclusion opened the way to assessing the model fit of our hypothetical model and comparing it to the measurement model. Furthermore, this exclusion ensured that the results are not driven by the inclusion of control variables. A graphical overview of the complete structural model appears in Appendix B. The control variables were added to the analyses in a second model. Subsequently, to increase the fit of the model including control variables, a third model was analyzed that includes only those control variables that had a statistically significant relationship with one of the key variables in Model 2 (job type and employee age with *PM quality*, employee education with *use of PMs*, and employee gender with *employee job performance*). Model 3 served as the basis for analyzing whether common method bias may still have played a role. For this purpose, a latent common method variable was added to the fourth model, which was estimated based on each first item of the scales *PM quality*, *use of PMs for monetary compensation*, *use of PMs for nonmonetary rewards*, *use of PMs for evaluation purposes*, and *employee job performance* (Podsakoff et al., 2003).

Finally, to test for H6–H8, we analyzed whether *PM quality* and the three uses of performance metrics are indeed mediators in the relationship between *PM participation* and *employee job performance*. We assess mediation using maximum likelihood bootstrapping, because this approach allows detection of opposite mediating effects in small samples (Shrout and Bolger, 2002). We estimated the indirect effects of *PM participation* on *employee job performance*, via (1) *PM quality* and *use of PMs for monetary compensation*, (2) *PM quality* and *use of PMs for nonmonetary rewards*, and (3) *PM quality* and *use of PMs for evaluation purposes*, respectively. We followed this approach for all four models.[11]

Moreover, a fully saturated model was run, in that direct effects were included between *PM participation* and *employee job performance*, between *PM quality* and *employee job performance*, and between *PM participation* and *use of PMs*. Such an analysis can reveal correlated omitted variables. If no correlated omitted variables are present, the three added direct relationships should be zero, since all variance should be explained by their mediators. In other words, if any of these relationships were significant in the fully saturated model, the analyses would suffer from omitted-variable bias.

We assessed model fit with several indices. First, we used chi-square and assessed the robustness of the chi-square model fit, using the Bollen-Stine bootstrap with 1999 bootstrap samples. Since the chi-square test of fit is sensitive to sample size, we also used other fit indices to check the model fit (Bentler, 1990). Schreiber et al. (2006) recommend CFI, TLI, and RMSEA for one-time analyses, where no comparisons of non-nested models are made. CFI and TLI are sufficient if they are around 0.95 or higher, and RMSEA should be around 0.08 or lower (Hu and Bentler, 1999).

4. Results

Table 2 shows the descriptive statistics of the survey items and the factor loadings of the measurement model. The confirmatory factor analysis shows that the measurement model fits very well ($\chi^2 = 209$, $df = 179$, $p = 0.06$, Bollen-Stine $p = 0.60$; CFI = 0.97; TLI = 0.97; RMSEA = 0.05). Table 3 displays the correlations between all constructs, including the control variables.[12]

The model fit of the structural model is also very good ($\chi^2 = 207$, $df = 180$, $p = 0.09$, Bollen-Stine $p = 0.64$; TLI = 0.97; CFI = 0.98; RMSEA = 0.04). Fig. 2 and Model 1 of Table 4 show the standardized regression weight estimates for the hypothesized relations.[13] The model provides strong to moderate support for H1, H2, and H5, that is, statistically significant relations are found. The null hypotheses of H3 and H4 could not be rejected, which means no significant relations are found. The significance levels are the same after doing a maximum likelihood bootstrap with 1999 samples (non-tabulated), except for the relationship between *use of PMs for monetary compensation* and *employee job performance* (H3), where its two-tailed significance is <0.05. Model 1 of Table 5 shows the unstandardized indirect effects to test H6–H8. Significant relations are found for H6 and H8, which rejects the null hypothesis of H6 and provides support for H8.

The results of adding all control variables are given in Model 2 of Tables 4 and 5. This model provides support for H1, H2, H5, and H8. Model 3 of Tables 4 and 5 is a simplified version of Model 2, which excludes all control variables and relationships that did not turn out to be significant in Model 2. Significant relations are found for H1, H2, H5, H6, and H8, which provides support for H1, H2, H5 and H8, and rejects the null hypothesis of H6. As mentioned earlier, we also statistically checked for common method bias. Model 4 of Tables 4 and 5 show the results of these analyses. The results again support H1, H2, H5, and H8. Furthermore, a fully saturated model was run for all models of Table 4 and none of the added direct effects was significant. All other relationships remained the same.

To summarize the above results, all four models support H1 (*PM participation–PM quality*), H2 (*PM quality–Use of PMs*), H5 (*Use of PMs for evaluation purposes–Employee job performance*), and H8 (the indirect effect between *PM participation* and *employee job performance* via *PM quality* and *using the PMs for evaluation purposes*). The null hypothesis of H6 (the negative indirect effect between *PM participation* and *Employee job performance* via *PM quality* and *Use of PMs for monetary compensation*) is rejected by Models 1 and 3, but not by Models 2 and 4. The null hypotheses of H3 (*Use of PMs for monetary compensation –Employee job performance*) and of the relationships with *Use of PMs for nonmonetary rewards* (H4 and H7) could not be rejected.

To check for robustness of the results, we also ran all analyses with other specifications of the model: (1) we used other specifications of the PM quality construct by also including the deleted items, and (2) we analyzed the models in which the second-order construct *use of PMs* was replaced by only the first-order constructs (hence *PM quality* was directly connected with *use of PMs for monetary compensation*, *use of PMs for nonmonetary rewards*, and *use of PMs for evaluation purposes*). The robustness checks all showed similar results. The measurement model of the latter model shows that—despite the finding that monetary compensation is relatively little used in Dutch firms (Jansen et al., 2009)—each of the variables has enough variance: for *use of PMs for monetary compensation* $SD^2 = 2.16$ ($p < 0.001$), for *use of PMs for nonmonetary rewards* $SD^2 = 2.56$ ($p < 0.001$), and for *use of PMs for evaluation purposes* $SD^2 = 0.82$ ($p < 0.001$). Furthermore, the model shows high correlations between the different types of use: for *use of PMs for monetary compensation–use of PMs for nonmonetary rewards* r = 0.64 ($p < 0.001$), for *use of PMs for monetary compensation–use of PMs for evaluation purposes* r = 0.71 ($p < 0.001$), and for *use of PMs for*

[11] To be able to perform the bootstraps necessary for estimating the indirect effects and their significance levels, missing variables were replaced by their means. The model fit indices and significance levels of the estimated model are similar to the ones of Table 4.

[12] These estimates differ less than 0.01 from the estimates of the model without control variables, and have the same significant levels.

[13] The factor loadings of the items are the same as in the measurement model (Table 1).

Table 2
Descriptive statistics and factor loadings of the measurement model.

Constructs and key words of underlying items	α	N	M	SD	Min	Max	Standardized factor loadings[a]	Unstandardized factor loadings[a]	R^2
PM participation	0.94								
1. Metrics' design		86	3.48	1.97	1	7	0.90	1.00***	0.80
2. Choice of data		86	3.58	1.86	1	7	0.88	0.93***	0.77
3. Ongoing modifications[b]		86	3.42	1.79	1	7	0.81	0.81***	0.65
4. Implementation		86	3.20	1.70	1	7	0.91	0.88***	0.83
5. Maintenance[b]		86	3.45	1.76	1	7	0.84	1.00***	0.70
PM quality	0.72								
6. Precision		86	4.48	1.55	1	7	0.46	1.64***	0.21
7. Accuracy		86	4.53	1.35	1	7	0.86	1.32***	0.74
8. Sensitivity		86	5.22	1.26	1	7	0.75	0.97***	0.56
9. Verifiability		86	5.10	1.28	1	7	0.54	0.83***	0.29
Use of PMs	0.79								
Use of PMs for monetary compensation	*0.84*						*0.82*	*1.00******	*0.67*
10. Salary increases		86	4.74	1.63	1	7	0.91	1.00***	0.84
11. Bonuses		86	4.84	1.84	1	7	0.80	0.99***	0.65
Use of PMs for nonmonetary rewards	*0.88*						*0.69*	*0.90******	*0.48*
12. Promotion		86	4.70	1.70	1	7	0.94	1.00***	0.87
13. Increase authority		86	4.80	1.61	1	7	0.84	0.86***	0.71
Use of PMs for evaluation purposes	*0.89*						*0.88*	*0.65******	*0.77*
14. Performance evaluation		86	5.28	1.06	1	7	0.85	1.00***	0.73
15. Official performance rating		86	5.37	1.21	1	7	0.96	1.29***	0.93
16. Periodic discussions		86	5.10	1.11	1	7	0.74	0.90***	0.55
Employee job performance	0.91								
17. Essential duties		86	6.03	0.99	2	7	0.75	1.00***	0.57
18. Required responsibilities		86	5.72	1.10	2	7	0.88	1.31***	0.78
19. Formal requirements		86	5.49	1.12	2	7	0.83	1.25***	0.68
20. Job description		86	5.50	1.00	2	7	0.85	1.15***	0.73
21. Obligatory aspects		86	5.35	1.17	2	7	0.78	1.23***	0.61

*** $p < 0.001$ (two-tailed).
[a] Only the estimated factor loadings are shown in the table. The loadings of the measures on all other constructs (than the one the measure is posited to indicate) are set to zero.
[b] The error terms of these two items were allowed to co-vary: r = 0.400***.

Table 3
Construct variances (in parentheses on the diagonal axis) and correlations.

	Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	PM participation	(3.09)***													
2	PM quality	0.25*	(0.51)**												
3	Use of PMs	0.22*	0.72***	(1.33)***											
4	Employee job performance	0.04	0.17	0.16	(0.54)***										
5	Gender[a]	0.03	0.12	0.14	0.29**	(0.20)***									
6	Education[b]	−0.12	0.03	0.24*	0.24**	0.13	(2.62)***								
7	Age	0.05	−0.20	−0.19	−0.07	−0.16	−0.16	(93.0)***							
8	Departmental tenure	0.04	−0.06	−0.14	−0.01	−0.19*	−0.17	0.30***	(37.7)***						
9	Ln size organization	0.01	−0.07	−0.04	−0.08	−0.08	0.18	0.14	0.17	(3.69)***					
10	Ln size department	0.07	−0.03	0.08	0.14	0.05	0.09	−0.04	0.34***	0.15	(1.19)***				
11	Job type[c]	−0.18	0.23*	0.29**	0.25**	0.24**	0.43***	−0.15	−0.10	0.01	−0.11	(0.57)***			
12	Length of implementation	−0.11	−0.12	−0.18	−0.07	−0.24**	−0.01	0.30***	0.51***	−0.05	−0.02	0.02	(14.9)***		
13	Information asymmetry	−0.08	0.03	0.03	0.12	−0.17	0.41***	0.19	0.08	0.29**	0.11	0.05	0.01	(1.76)***	
14	Delegation	0.15	0.01	0.04	0.16	−0.01	0.09	−0.04	0.05	0.26**	−0.05	0.06	−0.06	0.32**	(0.46)***

* $p < 0.10$ (two-tailed).
** $p < 0.05$ (two-tailed).
*** $p < 0.01$ (two-tailed).
[a] 1 = male; 2 = female.
[b] 1 = lower vocational; 2 = intermediate general; 3 = intermediate vocational; 4 = high-level high school; 5 = BSc; 6 = MSc or higher.
[c] 1 = operator; 2 = service employee; 3 = professional.

nonmonetary rewards–use of PMs for evaluation purposes r = 0.59 ($p < 0.001$). These results correspond to our assumption that the three types of use usually go hand in hand with each other. We allowed the error terms of the three latent variables to co-vary in the structural model, so that only their unique variance is used to predict employee job performance. The structural model shows a strong relationship between *PM quality* and all three ways of using the performance metrics: $\beta = 0.57$ ($p < 0.01$) for the relationship with *use of PMs for monetary compensation*, $\beta = 0.44$ ($p < 0.01$) for the relationship with *use of PMs for nonmonetary rewards*, and $\beta = 0.67$ ($p < 0.01$) for the relationship with *use of PMs for evaluation purposes* (the betas are similar for different specifications of the model).

5. Discussion

The development of suitable and valid operational performance metrics is challenging because the metrics need to include enough context-specific idiosyncrasies to be meaningful to the employees and managers who use the metrics. The main goal of this research

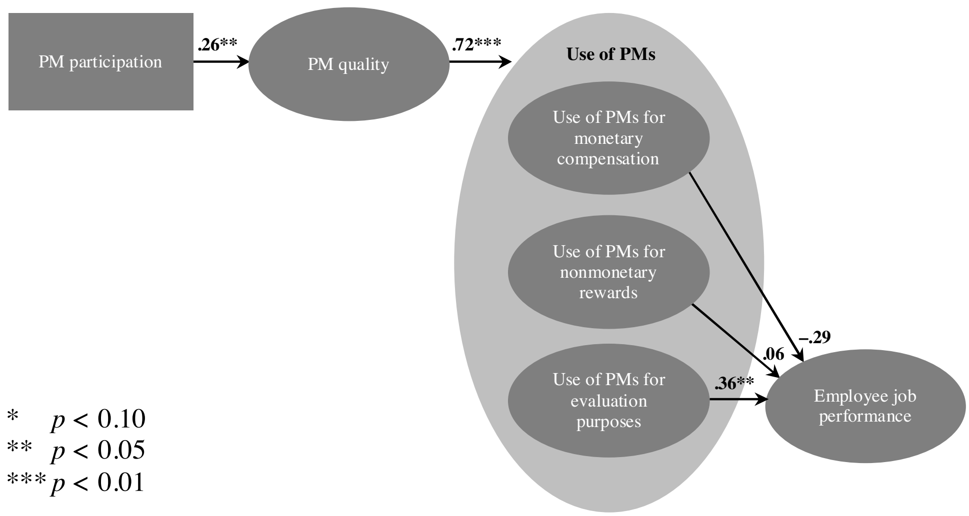


Fig. 2. Standardized solutions for the structural model (Model 1 of Table 4). For easy reference, the measurement model, which was also part of the analyses, is not shown in this figure but in Appendix B.

Table 4
Standardized coefficients and model fit of the structural model, with various specifications of control variables.

Hypothesis	Independent variable	Dependent variable	Model 1	Model 2	Model 3	Model 4
1 (+)	PM participation	PM quality	0.26**	0.34***	0.32***	0.32***
2 (+)	PM quality	Use of PMs	0.72***	0.72***	0.71***	0.70***
3 (0)	Use of PMs for monetary compensation	Employee job performance	−0.29	−0.23	−0.20	−0.19
4 (0)	Use of PMs for nonmonetary rewards	Employee job performance	0.06	0.00	0.02	0.01
5 (+)	Use of PMs for evaluation purposes	Employee job performance	0.36**	0.26*	0.28*	0.27*
Control	Job type[a]	PM quality		0.33*	0.27*	0.27*
Control	Age employee	PM quality		−0.23*	−0.19	−0.18
Control	Education employee[b]	Use of PMs		0.26*	0.18*	0.18*
Control	Gender employee	Employee job performance		0.23*	0.25*	0.25*
Model fit indices						
χ^2			207	918	295	289
df			180	654	256	251
p			0.09	0.00	0.05	0.05
CFI			0.98	0.87	0.97	0.97
TLI			0.97	0.83	0.96	0.96
RMSEA			0.04	0.07	0.04	0.04
Significant control variables included?			No	Yes	Yes	Yes
Other (non-significant) control variables included?			No	Yes	No	No
Latent common method variable included?			No	No	No	Yes

* $p < 0.10$ (one-tailed for H1, H2, H5, two-tailed for H3, H4, Control).
** $p < 0.05$ (one-tailed for H1, H2, H5, two-tailed for H3, H4, Control).
*** $p < 0.01$ (one-tailed for H1, H2, H5, two-tailed for H3, H4, Control).
N = 86.
[a] 1 = operator; 2 = service employee; 3 = professional.
[b] 1 = lower vocational; 2 = intermediate general; 3 = intermediate vocational; 4 = high-level high school; 5 = BSc; 6 = MSc or higher.
[c] 1 = male; 2 = female.

Table 5
Unstandardized indirect effects with various specifications of control variables.

Hyp.	Path					Model 1[a]	Model 2	Model 3	Model 4
H6	PM participation,	PM quality,	Use of PMs,	Use of PMs for monetary compensation,	Employee job performance	−0.018*	−0.019	−0.015*	−0.014
H7	PM participation,	PM quality,	Use of PMs,	Use of PMs for nonmonetary rewards,	Employee job performance	0.003	0.000	0.001	0.001
H8	PM participation,	PM quality,	Use of PMs,	Use of PMs for evaluation purposes,	Employee job performance	0.024**	0.026*	0.024**	0.023**
Significant control variables included?						No	Yes	Yes	Yes
Other (non-significant) control variables included?						No	Yes	No	No
Latent common method variable included?						No	No	No	Yes

*$p < 0.10$ (one-tailed for H8, two-tailed for H6, H7).
**$p < 0.05$ (one-tailed for H8, two-tailed for H6, H7).
***$p < 0.01$ (one-tailed for H8, two-tailed for H6, H7).
[a] The specifications of Model 1–4 are the same as in Table 4, with the following exceptions: to be able to estimate the indirect effects and their significance levels, in Models 2, 3 and 4 the 4 missing values for education were replaced by its mean and in Model 2 the 6 missing values for Ln Size of department and Ln Size of organization were additionally replaced by their means. Replacement is necessary because no missing values are allowed when using bootstrapping.

is to find out how managers can ensure that *PM participation* will have positive effects on employee job performance. We investigate the mediating mechanism between *PM participation* and *employee job performance*, and specifically look at variables from a management point of view—that is, *PM quality* as rated by the manager and *use of PIs for different incentive purposes*. The analyses show a statistically significant positive indirect effect from *PM participation* to *employee job performance* if the performance metrics are used for evaluation purposes. *Use of PMs for monetary compensation* or *nonmonetary rewards* do *not* seem to mediate the relationship between *PM participation* and *employee job performance*.

As the overall contribution, these findings on the indirect effects indicate that organizations might benefit from including operational employees in the development of performance metrics. In practical terms, the results advise managers to improve the quality of the performance metrics by involving employees in their development. Another piece of advice that emerges from the overall results is that managers should use these co-developed performance metrics for evaluative discussions about performance with their employees. With the help of such a set of performance measures, the manager and the employee could discuss the employee's performance, any problems the employee needs to address, favorable unexpected events, actions taken, improvement ideas for the next period, and so forth. The results did not show an effect on employee job performance of using the co-developed performance metrics for monetary compensation or nonmonetary rewards. Given the diverse findings of earlier research on the use of extrinsic rewards, managers are advised to first investigate what type of incentives work best for their operational employees within their organization, before using the performance measures as a basis for monetary compensation or nonmonetary rewards.

As a first additional contribution, the finding that managers are more positive about the quality of the performance metrics if employees have been involved in the design process (H1) expands what we know about how to develop valid and meaningful operational performance metrics. This research demonstrates that employee participation in developing these metrics is important to improving the metrics' quality in the eyes of the managers. Because including the knowledge of operational employees in the performance metrics is shown to be valuable, future research could investigate under what conditions managers are willing to start and sustain a PM participation *process*. This research could consider aspects such as the training of employees for such participation efforts, constraints and direction for the participation process, facilitation with focused methods, and the role of consultants and company specialists in areas such as accounting IT. Groen et al. (2012), for example, describe how a PM participation process was facilitated in a nuanced way. The process facilitator did not rely on expertise to "hand down" metrics, but helped employees develop their own ideas on new performance metrics and to make improvement initiatives work. The facilitator asked countless questions and offered follow-ups, built prototypes of proposed metrics, asked for continual feedback and resolutions, brought fresh ideas to the table, and constructively challenged employees' ideas.

Future research could also look at how to best reduce the fear that employees will abuse the process of co-development. Moreover, future research could determine other aspects that may improve performance measurement properties. Although not framed in this way, this study's results are consistent with studies using the concept of enabling formalization (e.g., Adler and Borys, 1996; Ahrens and Chapman, 2004; Chapman and Kihn, 2009; Englund and Gerdin, 2014; Mahama and Cheng, 2013; Patel, 2011; Tessier and Otley, 2012). Performance measurement may be enabling for both employees and managers, because even if they have diverging interests, both may benefit from performance metrics that have good measurement properties. Future research could therefore also investigate the effects of participation and metrics not only on employees being evaluated but also on the performance of managers in their roles as evaluators.

The second additional contribution of the study relates to the investigation of the differential effects of using the performance metrics for several purposes on employee job performance as predicted by self-determination theory. We differentiated between use of PMs for *monetary compensation*, *nonmonetary rewards*, and *evaluation purposes*, and we specifically looked at how much the unique variance of each of them (as compared to the other two) explains variance in employee job performance. It is important to use unique variance here, because this study is not about the job performance effects of *use of PMs* in general, but about the separate effects of the different uses. Of the three types of uses, we only find an increase in employee job performance if co-developed performance metrics are used for evaluation purposes (H5). We find no significant relation with performance when the PMs are used for monetary compensation or nonmonetary rewards. This finding is interesting, because in line with recent studies, it shows that the time-honored idea that the use of incentives has a positive effect on job performance is debatable.

Much more research on this incentives–performance link is necessary. Future research could, for example, concentrate on identifying and testing more moderators that may explain the many different findings with regard to this relationship between incentives and performance. Bonner and Sprinkle (2002) provide an overview of many person, task, and environmental variables that have been found to moderate the incentives–performance relation, and additional moderators are still being found (e.g., Fessler, 2003; Speklé and Verbeeten, 2014). A quick check, whereby we used the control variables of this study as interaction variables, showed a negative interaction effect of *use of PMs for monetary compensation* with departmental tenure of the employee and a positive interaction effect of *use of PMs for evaluation purposes* with age of the employee (both on *employee job performance*). Future research could test these and other potential interactions in other samples.

Furthermore, future research could be done in other countries or specific industries to obtain a broader empirical foundation and more nuanced understanding of the conditions under which our results hold. Incentive systems and their effects differ between countries (Jansen et al., 2009; Merchant et al., 2011). For example, a comparison of incentive practices within car dealerships in the Netherlands and the United States found a significant negative effect of using monetary compensation in Dutch firms, whereas this finding was insignificant in US firms (Jansen et al., 2009). Additionally, to increase insight into the exact workings of the effects of incentives on performance, future research could benefit from including measures of different types of motivation, such as the construct of psychological empowerment, used for instance by Hall (2008) and Marginson et al. (2014), or the motivation scale used by Kunz (2015). Eventually, such a stream of new studies will offer insights into how to best motivate the employee of the 21st century.

An important strength of this survey research is the use of multiple sources. By asking pairs consisting of an employee and the supervising manager to respond, we were able to assess all constructs at the relevant informant perspective and also to reduce common method bias. This study highlights the managers' views since they are the ones who decide whether to allow their employees to participate as well as how the performance metrics will be used. Finding these pairs of survey respondents was very difficult, especially given the strict criteria they had to meet. Many potential respondents stated their organization had plans to develop a performance measurement system, but had not yet been able to do so. Our effort resulted in a sample of 86 research pairs, which is smaller than we had hoped for. However, if the population is restricted in size, structural equation modeling may be applied to smaller samples (Barrett, 2007; Kline, 2011). Also, research has shown that this application does not increase the chance of false positives, although it does increase the chance of false negatives (Goodhue et al., 2012). On the positive side, we are able to detect only large effects. Furthermore, we performed many additional checks that show the robustness of the results.

We used a cross-sectional design in the current study because we wanted to test the hypotheses in a broad sample of organizations in the field. This design gives the present research external validity, but at the expense of the possibility to infer causality. Although the theory behind our model assumes a causal relationship, causality cannot be inferred before it has been tested with a longitudinal or experimental design (Van der Stede, 2014). For example, the relationship between *PM quality* and *Use of PMs* (H2) might also be a matter of self-serving bias—that is, rather than making more use of high-quality PMs, managers might perceive the PMs as being of higher quality to rationalize why they make more use of them. Furthermore, a reason we do not find a positive performance effect of the use of PMs for monetary compensation (H3) may be that managers would use monetary incentives with low-performing employees with the hope of motivating them to perform better. Starting with the outcome of the research reported here, we recommend the use of a quasi-experimental design in one or more large organizations.

To conclude, the results of the present study point to the need for operational employee participation in the development of performance metrics. This participation is likely to lead to operational performance metrics that managers perceive as having better measurement properties. Subsequent use of these metrics for evaluation purposes, rather than for distributing monetary compensation, may contribute to the improvement of employee job performance.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors thank the survey participants as well as everyone who helped us find respondents, including the following organizations: EVO, NGI Platform voor ICT-professionals, Noventum Service Management Consultants, TSM2Connect and Vereniging Logistiek Management. For their constructive questions and comments on earlier versions of this paper, we thank Matthew Hall, Jan Pfister, Theresa Libby, the journal reviewers, the workshop participants and anonymous reviewers of: the 6th Conference on Performance Measurement and Management Control; the 2012 American Accounting Association MAS Research and Case Conference; the 35th annual congress of the European Accounting Association; the 2012 Academy of Management Annual Meeting; the EurAM 2012 Annual Conference; the 2012 Limperg course Managerial Accounting; Erasmus University Rotterdam; University of Groningen; University of Twente; University of Amsterdam; University of Edinburgh; and University of Turku.

Appendix A. Items used in the scales.

Respondents rated all items on a seven-point fully anchored Likert scale: (1) totally disagree, (2) disagree, (3) moderately disagree, (4) neutral, (5) moderately agree, (6) agree, (7) totally agree.

*A.1. PM participation (Abernethy and Bouwens, 2005*Abernethy and Bouwens, 2005*; completed by the employee)*

I have/had an influence on:
1. How the performance metrics are designed
2. The choice of the data used as an input in the performance metrics
3. Ongoing modifications to the design of the performance metrics
4. The implementation of the performance metrics
5. The maintenance of the performance metrics

*A.2. PM quality (selection of Moers, 2006*Moers, 2006*; completed by the manager)*

6. The performance metrics measure only what my employee can actually influence
7. The performance metrics express accurately whether my employee functions well or not
8. If my employee performs well, it is directly reflected in the performance metrics
9. The performance metrics are objective and verifiable

*A.3. Use of PMs for monetary compensation (Moers, 2006*Moers, 2006*; completed by the manager)*

I attach very high importance to the performance metrics in:
10. Determining potential salary increases
11. Determining potential bonuses or extras

*A.4. Use of PMs for nonmonetary rewards (Moers, 2006*Moers, 2006*; completed by the manager)*

I attach very high importance to the performance metrics in:
12. Increasing my employee's chance of promotion
13. Increasing my employee's authority within the organization

*A.5. Use of PMs for evaluation purposes (selection of Moers, 2006*Moers, 2006*; completed by the manager)*

I attach very high importance to the performance metrics in:
14. The evaluation of my employee's performance
15. Officially rating my employee's performance
16. Periodic discussions with my employee

*A.6. Employee job performance (Podsakoff and MacKenzie, 1989*Podsakoff and MacKenzie, 1989*; completed by the manager)*

17. My employee always performs all essential duties
18. My employee always fulfills all responsibilities required by his/her job
19. My employee always meets all formal performance requirements of the job
20. My employee always completes all duties specified in his/her job description
21. My employee never neglects aspects of the job that he/she is obligated to perform

Appendix B. The structural model.

See Fig. B1.

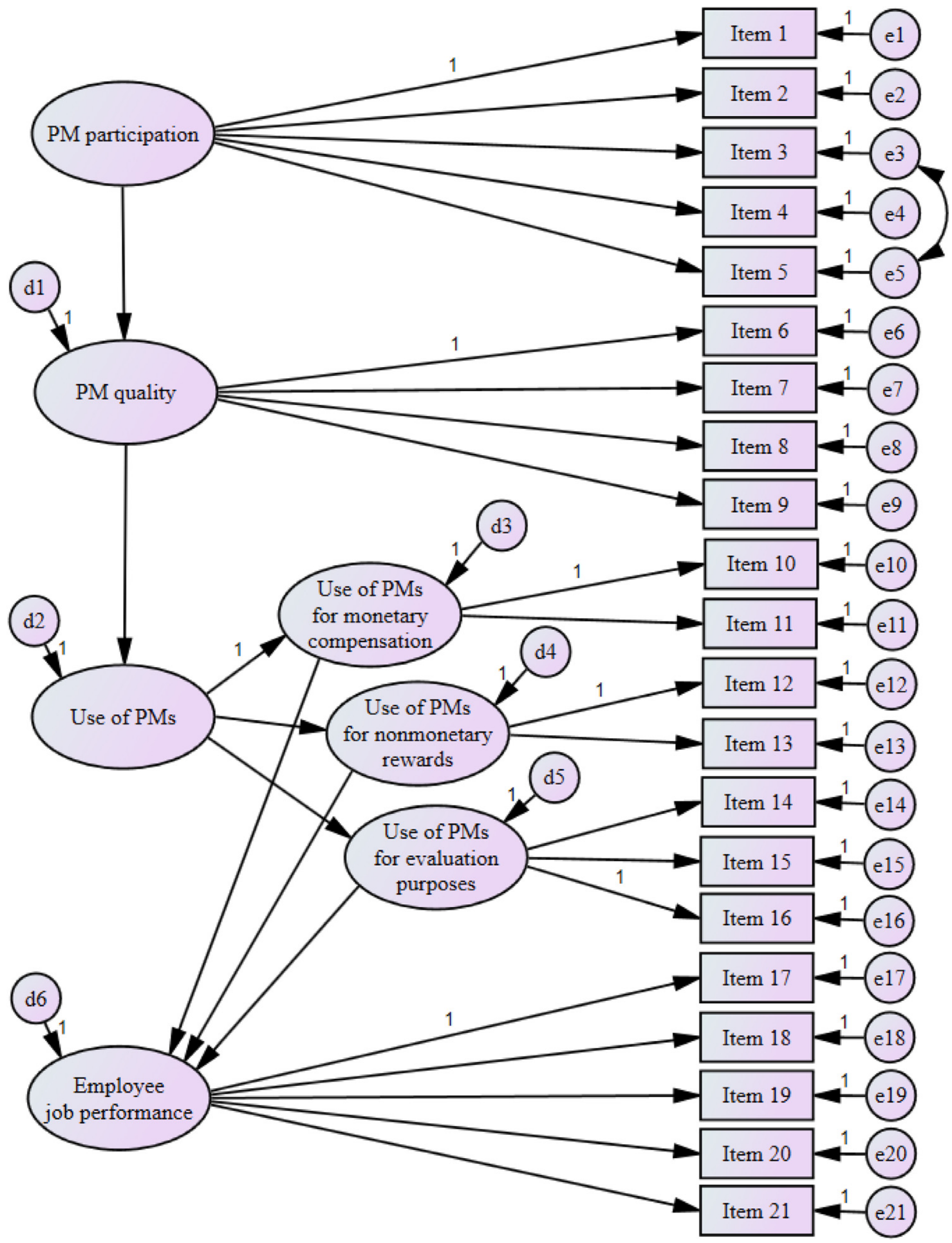


Fig. B1. Graphical representation of the structural model, including the second-order construct *Use of PMs* and its underlying first-order constructs.

References

Abernethy, M.A., Bouwens, J., 2005. Determinants of accounting innovation implementation. Abacus 41 (3), 217–240.

Adler, P.S., Borys, B., 1996. Two types of bureaucracy: enabling and coercive. Adm. Sci. Q. 41 (1), 61–89.

Ahrens, T.A., Chapman, C.S., 2004. Accounting for flexibility and efficiency: a field study of management control systems in a restaurant chain. Contemp. Account. Res. 21 (2), 271–301.

Ali, H., Davies, D.R., 2003. The effects of age, sex and tenure on the job performance of rubber tappers. J. Occup. Organ. Psychol. 76, 381–391.

Anderson, J.C., Gerbing, D.W., 1988. Structural equation modeling in practice: a review and recommended two-step approach. Psychol. Bull. 103 (3), 411–423.

Anderson, J.C., Gerbing, D.W., 1991. Predicting the performance of measures in a confirmatory factor analysis with a pretest of their substantive validities. J. Appl. Psychol. 76, 732–740.

Ashton, R.H., 1990. Pressure and performance in accounting decision settings: paradoxical effects of incentives, feedback, and justification. J. Account. Res. 28, 148–180.

Bénabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. Rev. Econ. Stud. 70, 489–520.

Baiman, S., Evans, J.H., 1983. Pre-decision information and participative management control systems. J. Account. Res. 21 (2), 371–394.

Banker, R.D., Datar, S.M., 1989. Sensitivity, precision, and linear aggregation of signals for performance evaluation. J. Account. Res. 27 (1), 21–39.

Banker, R.D., Lee, S.Y., Potter, G., Srinivasan, D., 2000. An empirical analysis of continuing improvements following the implementation of a performance-based compensation plan. J. Account. Econ. 30 (3), 315–350.

Barrett, P., 2007. Structural equation modelling: adjudging model fit. Pers. Individ. Differ. 42, 815–824.

Bentler, P.M., 1990. Comparative fit indexes in structural models. Psychol. Bull. 107, 238–246.

Bisbe, J., Otley, D., 2004. The effects of the interactive use of management control systems on product innovation. Account. Organ. Soc. 29, 709–737.

Bonner, S.E., Sprinkle, G.B., 2002. The effects of monetary incentives on effort and task performance: theories evidence, and a framework for research. Account. Organ. Soc. 27, 303–345.

Bonner, S.E., Hastie, R., Sprinkle, G.B., Young, S.M., 2000. A review of the effects of financial incentives on performance in laboratory tasks: implications for management accounting. J. Manag. Account. Res. 12, 19–64.

Bourne, M., Mills, J., Wilcox, M., Neely, A., Platts, K., 2000. Designing, implementing and updating performance measurement systems. Int. J. Oper. Prod. Manag. 20 (7), 754–771.
Brown III, J.L., Evans, J.H., Moser, D.V., 2009. Agency theory and participative budgeting experiments. J. Manag. Account. Res. 21, 317–345.
Burney, L.L., Henle, C.A., Widener, S.K., 2009. A path model examining the relations among strategic performance measurement system characteristics, organizational justice, and extra- and in-role performance. Account. Organ. Soc. 34, 305–321.
Burt, R.S., 1976. Interpretational confounding of unobserved variables in structural equation models. Sociol. Methods 5, 3–52.
Cameron, J., Pierce, W.D., 1994. Reinforcement, reward, and intrinsic motivation: a meta-analysis. Rev. Educ. Res. 64, 363–423.
Cameron, J., Banko, K.M., Pierce, W.D., 2001. Pervasive negative effects of rewards on intrinsic motivation: the myth continues. Behav. Analyst 24, 1–44.
Cameron, J., Pierce, W.D., Banko, K.M., Gear, A., 2005. Achievement-based rewards and intrinsic motivation: a test of cognitive mediators. J. Educ. Psychol. 97, 641–655.
Cameron, J., 2001. Negative effects of reward on intrinsic motivation—a limited phenomenon: comment on Deci, Koestner, and Ryan (2001). Rev. Educ. Res. 71, 29–42.
Chapman, C.S., Kihn, L., 2009. Information system integration, enabling control and performance. Account. Organ. Soc. 34, 151–169.
Chow, C.W., Cooper, J.C., Waller, W.S., 1988. Participative budgeting: effects of a truth-inducing pay scheme and information asymmetry on slack and performance. Account. Rev. 63 (1), 111–122.
Christ, M.H., Emett, S.A., Summer, S.L., Wood, D.A., 2012. The effects of preventive and detective controls on employee performance and motivation. Contemp. Account. Res. 29 (2), 432–452.
Christensen, J., 1982. The determination of performance standards and participation. J. Account. Res. 20 (2), 589–603.
Dalton, D.W., Buchheit, S., McMillan, J.J., 2014. Audit and tax career paths in public accounting: an analysis of student and professional perceptions. Account. Horiz. 28, 213–231.
De Baerdemaeker, J., Bruggeman, W., 2015. The impact of participation in stratigic planning on managers' creation of budgetary slack: the mediating role of autonomous motivation and affective organizational commitment. Manag. Account. Res. 29, 1–12.
De Haas, M., Algera, J.A., 2002. Demonstrating the effect of strategic dialogue: participation in designing the management control system. Manag. Account. Res. 13, 41–69.
Deci, E.L., Ryan, R.M., 2000. The what and why of goal pursuits: human needs and the self-determination of behavior. Psychol. Inq. 11 (4), 227–268.
Deci, E.L., Koestner, R., Ryan, R.M., 1999a. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol. Bull. 125, 627–668.
Deci, E.L., Koestner, R., Ryan, R.M., 1999b. The undermining effect is a reality after all—extrinsic rewards, task interest, and self-determination: reply to Eisenberger, Pierce, and Cameron (1999) and Lepper, Henderlong, and Gingras (1999). Pscyhol. Bull. 125, 692–700.
Deci, E.L., Koestner, R., Ryan, R.M., 2001a. Extrinsic rewards and intrinsic motivation in education: reconsidered once again. Rev. Educ. Res. 71, 1–27.
Deci, E.L., Ryan, R.M., Koestner, R., 2001b. The pervasive negative effects of rewards on intrinsic motivation: response to Cameron (2001). Rev. Educ. Res. 71, 43–51.
Derfuss, K., 2009. The relationship of budgetary participation and reliance on accounting performance measures with individual-level consequent variables: a meta-analysis. Eur. Account. Rev. 18 (2), 203–239.
Drake, A.R., Wong, J., Salter, S.B., 2007. Empowerment, motivation, and performance: examining the impact of feedback and incentives on nonmanagement employees. Behav. Res. Account. 19, 71–89.
Dunk, A.S., 1993. The effect of budget emphasis and information asymmetry on the relation between budgetary participation and slack. Account. Rev. 68 (2), 400–410.
Eisenberger, R., Cameron, J., 1996. Detrimental effects of reward: reality or myth? Am. Psychol. 51 (11), 1153–1166.
Eisenberger, R., Pierce, W.D., Cameron, J., 1999. Effects of reward on intrinsic motivation—negative, neutral, and positive: comment on Deci, Koestner, and Ryan (1999). Psychol. Bull. 125, 677–691.
Englund, H., Gerdin, J., 2015. Developing enabling performance measurement systems: on the interplay between numbers and operational knowledge. Eur. Account. Rev. 24 (2), 277–303.
Evans, J.R., 2004. An exploratory study of performance measurement systems and relationships with performance results. J. Oper. Manag. 22, 219–232.
Falk, A., Kosfeld, M., 2006. The hidden costs of control. Am. Econ. Rev. 96 (5), 1611–1630.
Fehr, E., Falk, A., 2002. Psychological foundations of incentives. Eur. Econ. Rev. 46, 687–724.
Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. J. Econ. Perspect. 14 (3), 159–181.
Feltham, G.A., Xie, J.L., 1994. Performance measure congruity and diversity in multi-task principal/agent relations. Account. Rev. 69 (3), 429–453.
Ferreira, A., Otley, D., 2009. The design and use of performance management systems: an extended framework for analysis. Manag. Account. Res. 20, 263–282.
Fessler, N.J., 2003. Experimental evidence on the links among monetary incentives, task attractiveness, and task performance. J. Manag. Account. Res. 15, 161–176.
Franco-Santos, M., Kennerly, 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. Manag. 27 (8), 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, 79–119.
Frantom, C., Green, K.E., Lam, T.C.M., 2002. Item grouping effects on invariance of attitude items. J. Appl. Meas. 3, 38–49.
Frey, B.S., Jegen, R., 2001. Motivation crowding theory. J. Econ. Surv. 15 (5), 589–611.
Frey, B.S., 1994. How intrinsic motivation is crowded out and in. Ration. Soc. 6, 334–352.
Gagné, M., Deci, E.L., 2005. Self-determination theory and work motivation. J. Organ. Behav. 26, 331–362.
Gibbs, M.J., Merchant, K.A., Van der Stede, W.A., Vargus, M.E., 2004. Determinants and effects of subjectivity in incentives. Account. Rev. 79 (2), 409–436.
Gibbs, M.J., Merchant, K.A., Van der Stede, W.A., Vargus, M.E., 2009. Performance measure properties and incentive system design. Ind. Relat. 48 (2), 237–258.
Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. J. Econ. Perspect. 25 (4), 1–21.
Goodhue, D.L., Lewis, W., Thompson, R., 2012. Does PLS have advantages for small sample size or non-normal data? MIS Q. 36, 981–1001.
Grafton, J., Lillis, A.M., Widener, S.K., 2010. The role of performance measurement and evaluation in building organizational capabilities and performance. Account. Organ. Soc. 35, 689–706.
Groen, B.A.C., Wouters, M.J.F., Wilderom, C.P.M., 2012. Why do employees take more initiative if they develop their own performance measures? A field study. Manag. Account. Res. 23 (2), 120–141.
Groen, B.A.C., Wilderom, C.P.M., Wouters, M.J.F., 2016. High job performance through co-developing performance measures with employees. Hum. Resour. Manag., http://dx.doi.org/10.1002/hrm.21762, in press.
Hak, T., Van der Veer, K., Jansen, H., 2008. The three-step test-interview (TSTI): an observation-based method for pretesting self-completion questionnaires. Surv. Res. Methods 2 (3), 143–150.
Hall, D.T., 1996. Protean careers of the 21 st century. Acad. Manag. Exec. 10, 8–16.
Hall, D.T., 2004. The protean career: a quarter-century journey. J. Vocat. Behav. 65, 1–13.
Hall, M., 2008. The effect of comprehensive performance measurement systems on goal clarity. Account. Organ. Soc. 33, 141–163.
Hall, M., 2010. Accounting information and managerial work. Account. Organ. Soc. 35, 301–315.
Hansen, S.C., Van der Stede, W.A., 2004. Multiple facets of budgeting: an exploratory analysis. Manag. Account. Res. 15 (4), 415–439.
Henri, J.-F., 2006. Management control systems and strategy: a resource-based perspective. Account. Organ. Soc. 31, 529–558.
Ho, J.L.Y., Wu, A., Wu, S.Y.C., 2014. Performance measures, consensus on strategy implementation, and performance: evidence from the operational-level of organizations. Account. Organ. Soc. 39, 38–58.
Holmås, T.H., Kjerstad, E., Luråsd, H., Straume, O.R., 2010. Does monetary punishment crowd out pro-social motivation? a natural experiment on hospital length of stay. J. Econ. Behav. Organ. 75, 261–267.
Holmström, B., 1979. Moral hazard and observability. Bell J. Econ. 10, 74–91.
Hopp, W., Iravani, S.M.R., Liu, F., 2009. Managing white-collar work: an operations-oriented survey. Prod. Oper. Manag. 18 (1), 1–32.
Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6 (1), 1–55.
Hunton, J.E., Gibson, D., 1999. Soliciting user-input during the development of an accounting information system: investigating the efficacy of group discussion. Account. Organ. Soc. 24, 597–618.
Jørgensen, B., Messner, M., 2010. Accounting and strategising: a case study from new product development. Account. Organ. Soc. 35, 184–204.
James, H.S., 2005. Why did you do that? an economic examination of the effect of extrinsic compensation on intrinsic motivation and performance. J. Econ. Psychol. 26, 549–566.
Jansen, E.P., Merchant, K.A., Van der Stede, W.A., 2009. National differences in incentive compensation practices: the differing roles of financial performance measurement in the United States and The Netherlands. Account. Organ. Soc. 34, 58–84.
Jansen, E.P., 2015. Participation, accounting and learning how to implement a new vision. Manag. Account. Res. 29, 45–60.
Jenkins Jr., G.D., Mitra, A., Gupta, N., Shaw, J.D., 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. J. Appl. Psychol. 83, 777–787.
Jensen, M.C., Meckling, W.H., 1992. Specific and general knowledge, and organizational structure. In: Werin, L., Wijkander, H. (Eds.), Contract Economics. Blackwell, Oxford UK, pp. 251–274.
Jordan, S., Messner, M., 2012. Enabling control and the problem of incomplete performance indicators. Account. Organ. Soc. 37, 544–564.
Kennerley, M., Neely, A., 2003. Measuring performance in a changing business environment. Int. J. Oper. Prod. Manag. 23 (2), 213–229.
Kim, Y.H., Sting, F.J., Loch, C.H., 2014. Top-down, bottom-up, or both? Toward an integrative perspective on operations strategy formation. J. Oper. Manag. 32 (7–8), 462–474.

Kleingeld, A., Van Tuijl, H., Algera, J., 2004. Participation in the design of performance management systems: a quasi-experimental field study. J. Organ. Behav. 25, 831–851.
Kline, R.B., 2011. Principles and practice of structural equation modelling, third ed. The Guilford Press, New York.
Koufteros, X., Babbar, S., Kaighobadi, M., 2009. A paradigm for examining second-order factor models employing structural equation modeling. Int. J. Prod. Econ. 120, 633–652.
Koufteros, X., Verghese, A., Lucianetti, L., 2014. The effect of performance measurement systems on firm performance: a cross-sectional and longitudinal study. J. Oper. Manag. 32, 313–336.
Kruis, A.-M., Widener, S.K., 2014. Managerial influence in performance measurement system design: a recipe for failure? Behav. Res. Account. 26 (2), 1–34.
Kunz, J., Linder, S., 2012. Organizational control and work effort: another look at the interplay of rewards and motivation. Eur. Account. Rev. 21 (3), 591–621.
Kunz, A.H., Pfaff, D., 2002. Agency theory, performance evaluation, and the hypothetical construct of intrinsic motivation. Account. Organ. Soc. 27, 275–295.
Kunz, J., 2015. Objectivity and subjectivity in performance evaluation and autonomous motivation: an exploratory study. Manag. Account. Res. 27, 27–46.
Lander, M.W., Koene, B.A.S., Linssen, S.N., 2013. Committed to professionalism: organizational responses of mid-tier accounting firms to conflicting institutional logics. Account. Organ. Soc. 38, 130–148.
Langevin, P., Mendoza, C., 2013. How can management control system fairness reduce managers' unethical behaviours? Eur. Manag. J. 31, 209–222.
Lau, C.M., Tan, S.L.C., 2005. The importance of procedural fairness in budgeting. Adv. Account. 21, 333–356.
Lau, C.M., Tan, S.L.C., 2006. The effects of procedural fairness and interpersonal trust on job tension in budgeting. Manag. Account. Res. 17, 171–186.
Lazear, E.P., 2000. Performance pay and productivity. Am. Econ. Rev. 90 (5), 1346–1361.
Li, P., Tang, G., 2009. Performance measurement design within its organisational context: evidence from China. Manag. Account. Res. 20, 193–207.
Lillis, A.M., 2002. Managing multiple dimensions of manufacturing performance—an exploratory study. Account. Organ. Soc. 27, 497–529.
Lourenço, S., 2016. Monetary incentive, feedback and recognition—complements or substitutes? evidence from a field experiment in a retail services company. Account. Rev. 91 (1), 279–297.
Mahama, H., Cheng, M.M., 2013. The effect of manager's enabling perceptions on costing system use, psychological empowerment, and task performance. Behav. Res. Account. 25 (1), 89–114.
Marginson, D., McAulay, L., Roush, M., Van Zijl, T., 2014. Examining a positive psychological role for performance measures. Manag. Account. Res. 25, 63–75.
McKinnon, S.M., Bruns Jr., W.J., 1992. The Information Mosaic. Havard Business School Press, Boston, MA.
Melnyk, S.A., Stewart, D.M., Swink, M., 2004. Metrics and performance measurement in operations management: dealing with the metrics maze. J. Oper. Manag. 22, 209–217.
Merchant, K.A., Van der Stede, W.A., Zheng, L., 2003. Disciplinary constraints on the advancement of knowledge: the case of organizational incentive systems. Account. Organ. Soc. 28, 251–286.
Merchant, K.A., Van der Stede, W.A., Lin, T.W., Yu, Z., 2011. Performance measurement and incentive compensation: an empirical analysis and comparison of Chinese and Western firms' practices. Eur. Account. Rev. 20 (4), 639–667.
Merchant, K.A., 1981. The design of the corporate budgeting system: influences on managerial behavior and performance. Account. Rev. 56 (4), 813–829.
Mills, P.K., Ungson, G.R., 2003. Reassessing the limits of structural empowerment: organizational constitution and trust as controls. Acad. Manag. Rev. 28 (1), 143–153.
Moers, F., 2006. Performance measure properties and delegation. Account. Rev. 81 (4), 897–924.
Neely, A.D., Richards, H., Mills, J., Platts, K., Bourne, M., 1997. Designing performance measures: a structured approach. Int. J. Oper. Prod. Manag. 17 (11), 1131–1152.
Neely, A.D., Bourne, M., Mills, J., Platts, K., Richards, H., 2002. Getting the Measure of Your Business. University Press, Cambridge.
Neu, D., Rahaman, A.S., Everett, J., 2014. Accounting and sweatshops: enabling coordination and control in low-price apparel production chains. Contemp. Account. Res. 31, 322–346.
Nouri, H., Parker, R.J., 1998. The relationship between budget participation and job performance: the roles of budget adequacy and organizational commitment. Account. Organ. Soc. 23 (5/6), 467–483.
Ortega, J., 2009. Employee discretion and performance pay. Account. Rev. 84 (2), 589–612.
Patel, P.C., 2011. Role of manufacturing flexibility in managing duality of formalization and environmental uncertainty in emerging firms. J. Per. Manag. 29, 143–162.
Peng, D.X., Lai, F., 2012. Using partial least squares in operations management research: a practical guideline and summary of past research. J. Oper. Manag. 30, 467–480.
Penno, M., 1984. Asymmetry of pre-decision information and managerial accounting. J. Account. Res. 22 (1), 177–191.
Podsakoff, P.M., MacKenzie, S.B., 1989. A Second Generation Measure of Organizational Citizenship Behavior. Unpublished Manuscript.
Podsakoff, P.M., MacKenzie, S.B., Lee, J., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88 (5), 879–903.
Poon, M., Pike, R., Tjosvold, D., 2001. Budget participation, goal interdependence and controversy: a study of a Chinese public utility. Manag. Account. Res. 12, 101–118.
Prendergast, C., 2008. Work incentives, motivation and identity. Am. Econ. Rev.: Pap. Proc. 98 (2), 201–205.
Presslee, A., Vance, T.W., Webb, R.A., 2013. The effects of reward type on employee goal setting, goal commitment, and performance. Account. Rev. 88 (5), 1805–1831.
Raschke, R.L., Krishen, A.S., Kachroo, P., 2014. Understanding the components of information privacy threats for location-based services. J. Inf. Syst. 28, 227–242.
Ryan, R.M., Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. 55 (1), 68–78.
Salganik, M.J., Heckathorn, D.D., 2004. Sampling and estimation in hidden populations using respondent-driven sampling. Sociol. Methodol. 34, 193–239.
Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A., King, J., 2006. Reporting structural equation modeling and confirmatory factor analysis results: a review. J. Educ. Res. 99 (6), 323–337.
Segers, J., Inceoglu, I., Vloeberghs, D., Bartram, D., Henderickx, E., 2008. Protean and boundaryless careers: a study on potential motivators. J. Vocat. Behav. 73, 212–230.
Shearer, B., 2004. Piece rates, fixed wages and incentives: evidence from a field experiment. Rev. Econ. Stud. 71 (2), 513–534.
Shields, J., Shields, M.D., 1998. Antecedents of participative budgeting. Account. Organ. Soc. 23 (1), 49–76.
Shrout, P.E., Bolger, N., 2002. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol. Methods 7 (4), 422–445.
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.
Sprinkle, G.B., 2003. Perspectives on experimental research in managerial accounting. Account. Organ. Soc. 28, 287–318.
Stajkovic, A.D., Luthans, F., 1997. A meta-analysis of the effects of organizational behavior modification on task performance, 1975–95. Acad. Manag. J. 40 (5), 1122–1149.
Stajkovic, A.D., Luthans, F., 2003. Behavioral management and task performance in organizations: conceptual background meta-analysis, and test of alternative models. Pers. Psychol. 56, 155–194.
Stone, D.N., Bryant, S.M., Wier, B., 2010. Why are financial incentive effects unreliable? An extension of self-determination theory. Behav. Res. Account. 22 (2), 105–132.
Tessier, S., Otley, D., 2012. From management controls to the management of controls. Account. Audit. Account J. 25 (5), 776–805.
Thibaut, J., Walker, L., 1975. Procedural justice: a psychological analysis. Erlbaum, Hillsdale, NJ.
Thomas, K.W., Tymon Jr., W.G., 1993. Empowerment Inventory. Xicom, Tuxedo, NY.
Van Veen-Dirks, P.M.G., 2010. Different uses of performance measures: the evaluation versus reward of production managers. Account Organ. Soc. 35, 141–164.
Van der Stede, W.A., Young, S.M., Chen, C.X., 2005. Assessing the quality of evidence in empirical management accounting research: the case of survey studies. Account. Organ. Soc. 30, 655–684.
Van der Stede, W.A., 2014. A manipulationist view of causality in cross-sectional survey research. Account. Organ. Soc. 39, 567–574.
Webb, A., Jeffrey, S.A., Schulz, A., 2010. Factors affecting goal difficulty and performance when employees select their own performance goals: evidence from the field. J. Manag. Account Res. 22 (1), 209–232.
Wentzel, K., 2002. The influence of fairness perceptions and goal commitment on managers' performance in a budget setting. Behav. Res. Account. 14, 247–271.
Westland, J.C., 2010. Lower bounds on sample size in structural equation modeling. Electron. Commer. Res. Appl. 9, 476–487.
Widener, S.K., Shackell, M.B., Demers, E.A., 2008. The juxtaposition of social surveillance controls with traditional organizational design components. Contemp. Account. Res. 25 (2), 605–638.
Widener, S.K., 2007. An empirical analysis of the levers of control framework. Account. Organ. Soc. 32, 757–788.
Williams, L.J., Anderson, S.E., 1991. Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. J. Manag. 17 (3), 601–617.
Williamson, M.G., 2008. The effects of expanding employees decision making on contributions to firm value in an informal reward environment. Contemp. Account. Res. 25 (4), 1183–1209.
Wong-On-Wing, B., Guo, L., Lui, G., 2010. Intrinsic and extrinsic motivation and participation in budgeting: antecedents and consequences. Behav. Res. Account. 22 (2), 133–153.
Wouters, M.J.F., Roijmans, D., 2011. Using prototypes to induce experimentation and knowledge integration in the development of enabling accounting information. Contemp. Account. Res. 28 (2), 708–736.
Wouters, M.J.F., Wilderom, C.P.M., 2008. Developing performance-measurement systems as enabling formalization: a longitudinal field study of a logistics department. Account. Organ. Soc. 33, 488–516.