

University of Groningen

The effects of enabling versus coercive performance measurement systems on procedural fairness and red tape

Van Veen-Dirks, Paula; Leliveld, Marijke; Kaufmann, Wesley

Published in:
Journal of Management Control

DOI:
[10.1007/s00187-021-00316-5](https://doi.org/10.1007/s00187-021-00316-5)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Van Veen-Dirks, P., Leliveld, M., & Kaufmann, W. (2021). The effects of enabling versus coercive performance measurement systems on procedural fairness and red tape. *Journal of Management Control*, 32, 269-294. <https://doi.org/10.1007/s00187-021-00316-5>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



The effect of enabling versus coercive performance measurement systems on procedural fairness and red tape

Paula M. G. van Veen-Dirks¹ · Marijke C. Leliveld¹ · Wesley Kaufmann²

Accepted: 17 January 2021 / Published online: 13 March 2021
© The Author(s) 2021

Abstract

In this study, we investigate the effects of an enabling versus a coercive performance measurement system on how employees perceive the procedural quality of such systems. In particular, we examine the design characteristics and the development process of performance measurement systems. We hypothesize that an enabling design and an enabling development process, as compared to a coercive design and a coercive development process, lead to perceptions of greater procedural fairness and less red tape. To test our hypotheses, we conduct an experiment with two different samples (a student laboratory sample and an online sample). In general, our results indicate that an enabling performance measurement system design and an enabling system development process both independently increase procedural fairness and decrease red tape. These findings imply that organizations interested in improving the procedural quality of their performance measurement system should focus on designing and developing a system that is enabling rather than coercive.

Keywords Enabling · Coercive · Procedural fairness · Red tape · Performance measurement system

✉ Paula M. G. van Veen-Dirks
p.van.veen-dirks@rug.nl

Marijke C. Leliveld
m.c.leliveld@rug.nl

Wesley Kaufmann
w.kaufmann@uvt.nl

¹ Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

² Tilburg University, Tilburg, The Netherlands

1 Introduction

Management control and performance measurement systems are not only used in organizations to ensure that employees' behavior is consistent with organizational objectives and strategies, but also to help employees in doing their job, in searching for opportunities, and in solving problems (Mundy, 2010; Wouters & Wilderom, 2008). While the literature often assumes that the latter type—an enabling system—leads to a positive attitude, and the first type—a coercive system—leads to a negative attitude, this is not straightforward: the picture is usually more complex (Väisänen et al., 2020). This study adds to the current understanding of enabling versus coercive performance measurement and control by separating the concept of a system being either enabling or coercive from the assessment of the quality of the system.

In essence, Tessier and Otley (2012) argue that using labels such as enabling and coercive is not appropriate where one is also describing a positive or negative assessment of the system quality. They further argue that this distinction is particularly important since a management control and performance measurement system may be designed with an intent defined by management that fails to correspond with *employees' perceptions* of the system. That is, while a performance measurement system may be developed and designed in a way that is considered enabling by management, employees working with the system may perceive it differently. Acknowledging the importance of this employee perspective (e.g. Van der Kolk & Kaufmann, 2018; Van der Kolk et al., 2019), our research question is as follows: How are employees' perceptions of the procedural quality of a performance measurement system affected by the enabling or coercive orientation of such a system?

By addressing this question, we aim to enhance understanding of how individuals respond to different types of performance measurement in general, and to enabling and coercive systems in particular. Specifically, we aim to shed light on the effects of enabling or coercive performance measurement systems on employees' perceptions of a system's procedural quality through the lenses of procedural fairness (e.g. Colquitt et al., 2001; Leventhal, 1980) and of red tape (e.g. Bozeman, 1993). Note that while both procedural fairness and red tape provide an indication of employees' perceptions of procedural quality, the concepts capture distinct procedural quality dimensions (fairness versus the effectiveness, necessity, and burdensomeness of organizational procedures).

The term enabling generally refers to procedures that capture organizational memory, codify best practices, and help employees to deal more effectively with the inevitable contingencies associated with work processes. In contrast, coercive refers to forcing compliance, leaving little room for deviation from the rules and procedures in place, and reducing the role of employee commitment. In the literature on enabling and coercive systems, two dimensions have been distinguished: design characteristics of the system, and the development process of the system (e.g. Adler & Borys, 1996; Wouters & Wilderom, 2008). A performance measurement system's design is considered as enabling if it has four characteristics

(repair, flexibility, internal transparency, and global transparency). An enabling development process involves users in the design and implementation of a performance measurement system. In developing our hypotheses, we draw on the procedural fairness and red tape literatures to argue that both an enabling design and an enabling development process, relative to a coercive design and a coercive development process, lead to higher levels of perceived procedural quality of the performance measurement system as captured by procedural fairness and red tape.

We empirically test our hypotheses using a survey experiment. The experiment entails a vignette of a fictitious performance measurement system at a large business school. Participants are randomly assigned to one of four conditions that differ on the design characteristics (enabling vs. coercive) and development processes (enabling vs. coercive). The experiment is conducted in a laboratory setting with student participants and empirically replicated online with a sample using Amazon's Mechanical Turk (MTurk) and citizens as participants. Both samples allow us to test the relationships between the design characteristics and development processes of a performance measurement system and both perceived procedural fairness and perceived red tape. Having these two distinct samples help us to increase the validity of the study and to generalize the results beyond a student population.

Our study offers two main contributions to the literature on enabling and coercive performance measurement and management control systems. First, we contribute to the literature by acknowledging the conceptual distinction between enabling and coercive performance measurement systems and the perceived procedural quality of such systems. We argue that this enables a better understanding of how employees respond to performance measurement systems and management control systems (see also Van der Kolk & Kaufmann, 2018; Van der Kolk et al., 2019). Furthermore, our experimental design enables causal claims to be made regarding the relationship between enabling and coercive performance measurement systems and the perceived procedural quality of these systems.

Second, previous studies on management control and performance measurement have looked at a single dimension of enabling versus coercive systems (Väisänen et al., 2020): some studies have focused on the design characteristics of the system (Ahrens & Chapman, 2004; Jordan & Messner, 2012), while others were primarily interested in the process of developing the system (Wouters & Wilderom, 2008). In this paper, we add to this stream of research by systematically examining the effects of the *design characteristics* and the *development process* of performance measurement systems on employees' attitudes toward the system. This is particularly useful given that in previous studies, which were predominantly qualitative in nature, the potential overlap between performance measurement design and the development process in real-life settings could have made it difficult to clearly distinguish between design and process.

The structure of the paper is as follows. First, we develop our hypotheses linking enabling and coercive performance measurement to procedural fairness and red tape. We then present our experimental design and methods. The subsequent section then reports our results and we end with a concluding section that discusses the findings.

2 Enabling and coercive performance measurement

In this section, we build on the framework by Adler and Borys (1996) in which they make a distinction between enabling and coercive types of formalization. Since performance measurement can be conceived as a form of formalization, this framework can be used to contrast these two types of performance measurement. While Adler and Borys (1996, p. 66) propose that “employees’ attitudes to the system depend on the type of formalization with which they are confronted”, Tessier and Otley (2012) emphasize that the orientation of the system, as enabling or coercive, needs to be separated from the assessment of the quality of the system to increase conceptual clarity and to make it possible to empirically examine the relationship between these two concepts. Building on these insights, we develop hypotheses about the relationship between the type of performance measurement system—enabling or coercive—and the employees’ attitudes to these systems as captured by procedural fairness and red tape.

Although the enabling and coercive formalization framework was introduced in the context of equipment technology (Adler & Borys, 1996), scholars have applied the framework to various other settings including work teams (Proenca 2010), school structures (Hoy & Sweetland, 2001), and new product development (Jørgensen & Messner, 2009). Several studies have applied the framework to management control systems (Ahrens & Chapman, 2004) and in particular to performance measurement systems (Jordan & Messner, 2012; Wouters & Wilderom, 2008). These studies consider management control and performance measurement to be a form of formalization. While this emerging line of research on enabling and coercive systems tends to adopt the perspective of those in lower organizational positions that are subject to the system (Bisbe et al., 2019), the effect of an enabling, versus a coercive, system on employees’ attitudes toward the system has been left largely unexplored.

Several scholars have investigated the characteristics of management control systems that make people experience them as more or less fair. These studies have focused on a range of characteristics including subjective versus objective systems (Bellavance et al., 2013), diversity of performance measures and a focus on outcome versus effort in performance measures (Hartmann & Slapničar, 2012), bonus payments (Vouřem et al., 2016), and participation in goal setting (Groen 2018). Some also highlight the importance of the management control system for procedural fairness and, for example, emphasize that an important dysfunctional consequence of a management control system can be that the system is not experienced as fair by the employees (see e.g. Cugueró-Escofet & Rosanas, 2013). These authors further emphasize the importance of fairness by referring to the work of Folger and Cropanzano (1998): “when individuals perceive a lack of fairness, their morale declines, they become more likely to leave their jobs, and may even retaliate against the organization. Fair treatment, by contrast, breeds commitment intentions to remain on the job and helpful citizen behavior that go beyond the call of formal duties” (Folger & Cropanzano, 1998 preface xii; as referred to in Cugueró-Escofet & Rosanas, 2013).

While these studies provide relevant insights into how management control influences procedural fairness, we focus on characteristics of the management

control system previously not studied. As such, our research complements these studies by examining the effects of enabling and coercive performance measurement on procedural fairness. In addition, while the literature often refers to several normative procedural justice principles that can usually be traced back to the work of Leventhal (1980), Hartmann and Slapničar (2012, p. 17) comment that “these are not observable or designable characteristics of the performance evaluation themselves”. As a consequence, there is still only limited understanding of what the actual design and development process of the performance measurement system should be to achieve procedural fairness and, more broadly, a positive attitude toward the performance measurement system. Unlike procedural fairness, which has been studied in the management accounting and control literature (e.g. Groen 2018; Hartmann & Slapničar, 2012), the red tape concept that we include as another outcome variable in our study has not been well addressed.

Although the concepts of procedural fairness and red tape are related, they capture different elements of quality. Procedural fairness, i.e. the perceptions that decision-making procedures are fair (e.g. Lind & Tyler, 1988), is known to influence a wide range of citizen and employee attitudes and behaviors (Colquitt et al., 2001). For example, previous research has shown that procedural fairness improves satisfaction and the acceptance of unfavorable procedural outcomes (e.g. Dolan et al., 2007). Red tape refers to perceptions of unnecessarily burdensome organizational rules (Bozeman & Feeney, 2011; Kaufmann & Feeney, 2014). Here, organizational rules that lack effectiveness and efficiency, and that have excessive compliance burdens (Borri, 2016) have been found to unfavorably impact organizational performance (Brewer & Walker, 2010), as well as employee alienation (DeHart-Davis & Pandey, 2005) and satisfaction (Kaufmann & Tummers, 2017). Both procedural fairness and red tape relate to how employees perceive the procedures in place, and particularly to employees’ attitudes to the performance measurement system they are subject to. As such, we study the effects of an enabling, versus a coercive, performance measurement system on both perceived procedural fairness and perceived red tape.

2.1 Performance measurement systems and procedural fairness

In the framework by Adler and Borys (1996), an enabling design is characterized by four generic characteristics, namely repair, flexibility, internal transparency, and global transparency. A design lacking these formalization characteristics is viewed as coercive (Adler & Borys, 1996). *Repair* refers to the ease with which employees can fix the formal system in the event of a problem, while *flexibility* has to do with how much leeway exists for employees to deviate from the formal system. *Internal transparency* relates to the extent to which employees are able to understand the logic of the formal system itself, and *global transparency* refers to how much insight is provided into how the system fits into a broader organizational context. Each of these characteristics is expected to contribute to an enabling performance measurement system’s effect on employees’ perceptions of procedural quality, as captured by procedural fairness.

Repair means that the performance measurement system facilitates employees to respond to inevitable contingencies in the workplace and that employees are implicitly trusted to actively provide suggestions for improvements (Ahrens & Chapman, 2004) since some freedom for the employees is needed to determine the appropriate course of action in a set of given circumstances (Chapman & Kihn, 2009). In a coercive logic, there are no repair options. Under this logic, deviations from the formal performance measurement procedures are seen as suspect since the main purpose of coercive formalization is to ensure that employees' actions are compliant. Furthermore, coercive systems do not allow employees to repair procedural breakdowns themselves but instead require help from experts (Jorgenson & Messner, 2009).

Performance measurement that excludes repair opportunities limits employees' autonomy and discretion, making it more likely that such a procedure is perceived as unfair. Evidently, the opposite dynamic holds for enabling procedures that include repair options. Providing repair opportunities makes it possible for employees to resolve issues in the procedures that follow from unforeseen circumstances (Ahrens & Chapman, 2004), thus making it more likely that employees consider the procedures as fair, i.e. that perceived procedural fairness is high. This line of reasoning is seen in the procedural fairness literature, where correctability is viewed as a determinant of procedural fairness (Cugueró-Escofet & Rosanas, 2013; Groen, 2018; Leventhal, 1980; Leventhal et al., 1980).

Flexibility is concerned with the extent to which formalization provides options for employees to modify the system to best suit their own needs (Adler & Borys, 1996). Flexibility gives employees choices when addressing problems (Dowling & Leech, 2014). A flexible system assumes that deviations from procedures can provide learning opportunities and can make it easier to make decisions when confronted with emerging events (Chapman & Kihn, 2011; Wouters & Wilderom, 2008). In contrast, a coercive procedure forces employees to follow a specific sequence of steps. Deviations are not allowed unless the employee first gains approval from the supervisor (Adler & Borys, 1996). Although there is a risk that too much flexibility comes at the expense of consistency (Hartmann & Slapnicar, 2012)—another important condition for procedural fairness (Leventhal et al., 1980)—relevant information needs to be taken into account as much as possible to increase accuracy (see also Leventhal et al., 1980). As such, a lack of flexibility in formalization, and similarly in performance measurement, is likely to lead to negative outcomes, including low perceived procedural fairness. Ensure that decisions are based on as much good information and informed opinion as possible.

Internal transparency relates to the visibility of the system's internal workings and the extent to which the logic behind the system is understood by employees. A thorough understanding is also necessary to make the repair characteristic of the system useful (Chapman & Kihn, 2009). In contrast, coercive procedures are "formulated as lists of flat assertions and duties" (Adler & Borys, 1996, p. 72) and are designed to help supervisors rather than employees. In a coercive logic, it does not matter whether employees properly understand the system that they are working with, as long as they adhere strictly to the specified formal procedures. However, if the functioning of the system is known, employees will view it in a more positive light (DeHart-Davis, 2009), and the procedures are likely to be

perceived as more consistent and free from bias, which are important conditions for procedural fairness (Colquitt & Jackson, 2006; Leventhal et al., 1980). As a result, perceived procedural fairness is likely to be higher.

Finally, *global transparency* refers to the comprehensibility of the overall context within which employees are working. In other words, global transparency provides an understanding of how an employee's tasks fit into the organization as a whole (Ahrens & Chapman, 2004). If employees are allowed to adjust the system to unexpected contingencies, it is important that they understand how their work fits the larger organizational strategy and agenda (Chapman & Kihn, 2009). An enabling performance measurement system provides employees with a wide range of information that goes beyond their own domain (Wouters & Wilderom, 2008), thus helping employees to interact with the broader organizational context (Adler & Borys, 1996). In a coercive system, tasks are partitioned, and employees are only given access to information about the specific areas for which they are personally responsible. In such a system, global transparency is "a risk to be minimized" (Adler & Borys, 1996, p. 73) and employee compliance with organizational policies is forced without facilitating understanding (Dowling & Leech, 2014). DeHart-Davis (2009, 373) finds that "learning a rule's purpose could transform it from being perceived as a bad rule into a good rule", which can be applied in a performance measurement context as well. Against this background, a performance measurement system that is globally transparent is expected to lead to higher procedural fairness.

To summarize, we argue that performance measurement systems with enabling design characteristics lead to higher procedural fairness than coercive performance measurement systems that lack these design characteristics. This leads to our first hypothesis:

H1a: An enabling performance measurement system design leads to a higher level of perceived procedural fairness than a coercive design

An enabling orientation of a performance measurement system is not just reflected in its design, but also in its development process. Generally speaking, the development process of a performance measurement system can be considered as a mutually constituted and iterative process (Wouters, 2009; Wouters & Wilderom, 2008). An enabling development process encourages employee involvement, while a coercive process excludes such input from employees. Two particularly relevant elements of employee involvement in the formalization development process identified in the literature are building on employee experience and experimentation.

Wouters and Wilderom (2008, p. 493) argue that an experience-based development process "involves the identification, appreciation, documentation, evaluation, and consolidation of existing local knowledge and experience" and enhances the enabling nature of the system in place. It is particularly relevant that providing employees with the opportunity to test, review, and refine formal procedures also results in more positive attitudes towards the system because employees

will have a better understanding of the procedures' intended meaning (Wouters & Wilderom, 2008). As such, experimentation will favorably affect assessments of procedural fairness. For example, Wouters and Roijmans (2011) showed how experimentation during the process of developing a performance measurement system could help managers to integrate accounting information in an enabling way.

Part of an enabling development process is providing individuals with opportunities to participate in a decision-making process by giving them voice. Previous studies convincingly show that involving employees and giving them voice increases the perceived fairness of procedures (e.g. Colquitt et al., 2001; Groen, 2018, Lind & Tyler, 1988). The enabling process also ensures that the opinions of various groups affected by the decision are taken into account, which is an important condition for procedures to be perceived as fair (Leventhal et al., 1980). We therefore expect adopting an enabling development process for a performance measurement system to have a positive impact on perceived procedural fairness.

The above rationales are in line with other work showing that an enabling development process can lead to positive outcomes, improving organizational learning and employee motivation in particular (Groen et al., 2012; Jorgenson & Messner, 2009; Wouters, 2009; Wouters & Wilderom, 2008). This leads to our next hypothesis.

H1b: An enabling performance measurement system development process leads to a higher level of perceived procedural fairness than a coercive development process

2.1.1 Performance measurement systems and red tape

The four generic design characteristics of enabling formalization (repair, flexibility, internal transparency, and global transparency) also play a part in determining perceptions of red tape.

First, a lack of repair options in procedures will inevitably bring the work process to a halt (Adler & Borys, 1996) and subsequently result in administrative delay which is a key red tape indicator, especially in the context of a significant administrative burden (Bozeman & Feeney, 2011). A lack of repair options in a performance measurement system may thus cause procedural delays that result in a higher level of perceived red tape. In contrast, if employees have the opportunity to repair procedural breakdowns themselves, which is part of an enabling performance measurement system, perceived red tape is likely to be minimized.

A flexible performance measurement system allows deviations from the procedures. If a system does not allow employees to ignore, or bypass, a specific sequence of steps it will be perceived as over-controlling (Adler, 1999), and the associated inefficiencies will likely result in a negative assessment of procedural quality. In this light, DeHart-Davis (2009) notes that optimally-controlling rules are flexible, whereas over-controlling rules are inflexible. A lack of flexibility, as observed in coercive systems, can be linked to managerial and political control (Bozeman, 1993), and, in turn, to perceived red tape (Bozeman & Anderson, 2016).

A performance measurement system design that lacks internal transparency is more likely to be perceived as red tape, as perceptions of red tape are usually not the result of the existence of rules and procedures per se, but of their perceived ineffectiveness (Kaufmann and Feeney 2014). This line of reasoning also applies to performance measurement: if employees feel that a performance measurement procedure involves unnecessary compliance burdens, and could be redesigned to increase its effectiveness, it is likely to be associated with higher levels of red tape.

Finally, global transparency, which relates to the understandability of a rule's objectives, can be considered an important determinant of an individual's assessment of rule effectiveness. With a lack of global transparency, a failure on the part of the rule-maker to communicate the rule's objective to the people expected to comply with it may result in a higher level of perceived red tape (Bozeman, 2000; Bozeman & Feeney, 2011).

To summarize, the presence of enabling formalization design characteristics is hypothesized to lead to less perceived red tape.

H2a: An enabling performance measurement system design leads to a lower level of perceived red tape than a coercive design

With an enabling development process, employees are involved in the development of the performance measurement system. Since they have the opportunity to experiment with the procedures that are part of the system, they usually have a better understanding of the procedures' intended meaning (Wouters & Wilderom, 2008). This results in more positive attitudes towards the system and therefore in less perceived red tape. An enabling development process is also expected to lead to a lower level of perceived red tape because such a process can help prevent what is known in the red tape literature as rule-inception red tape, i.e. rules that are dysfunctional at their origin (Bozeman 1993, 2000). Rule-inception red tape is caused by various reasons, such as a misunderstanding of the problem at hand by the people designing the rules, or rules that are introduced by managers to compensate for a lack of control (Bozeman & Feeney, 2011). In an enabling process of developing a performance measurement system, employees can use their work experiences to identify salient organizational problems, and shortcomings in performance measurement procedures. As such, an enabling development process can help prevent rule-inception red tape. Based on the above, we expect an enabling development process to lead to a lower level of perceived red tape than a coercive development process. This leads to our final hypothesis:

H2b: An enabling performance measurement system development process leads to a lower level of perceived red tape than a coercive development process

Note that our experiment design enables us to disentangle the effects of design characteristics of the performance measurement system and of the development process of this system. As a consequence, we can also assess whether one of these factors moderates the other. Insights from motivational psychology, and in particular

from self-determination theory, suggest that a management control system which is perceived as controlling—as an extrinsic motivator—may have a negative effect on employees' autonomous motivation (Van der Kolk et al., 2019; Wong-on-Wing et al., 2010). Depending on the extent to which such an extrinsic motivator is internalized, it may also have a role in fostering autonomous motivation and positive outcomes (Deci & Ryan, 2000; Pfister & Lukka, 2019). Given that an enabling development process leads to an increased understanding of the underlying purpose of the performance measurement system as a result of the involvement of the employees (Wouters & Wilderom, 2008), employees may also internalize the value of the performance measurement system. The negative effect of systems with a coercive design on the employees' attitude toward the system (addressed in H1a and H2a) would thus be attenuated through an enabling development process because the conditions would be shaped for employees to internalize the external motivator. With such an enabling development process, a coercive design is thus likely to lead to a higher level of perceived procedural fairness and a lower level of perceived red tape than with a coercive process.

In contrast, however, one could argue that the increased understanding of the performance measurement system increases the appreciation of the enabling design characteristics. Moreover, a performance measurement system that is developed with an enabling process is likely to reflect the knowledge and concerns of the employees that were involved. The rules and procedures are thus more likely to be based on a good understanding of the organizational setting and the problems at hand. One might thus expect that the enabling design characteristics could be used to their full potential. For example, the repair options will be such that employees are in a good situation to fix any breakdowns themselves. One would then expect stronger effects if there were both an enabling design and an enabling process. In light of the above, while it may seem that an enabling development process can protect against the negative effects of a coercive performance measurement system design, the opposite may occur. Given these uncertainties, the interaction effects of development process and design are examined in an exploratory way rather than by testing hypotheses.

3 Method and data

3.1 Experimental design

We used a 2 (Characteristics: Enabling vs Coercive) \times 2 (Development: Enabling vs Coercive) between-subjects factorial design to test our hypotheses. Participants were asked to see themselves as a teacher working at a large business school and were told that, within this school, a Performance Measurement System (PMS) was used to monitor teaching performance. To manipulate the design characteristics, participants allocated to the *Enabling Design Characteristics* condition received information that the purpose of the system was to become a better lecturer (global transparency). Moreover, they read that performance goals were clearly defined, and that they were allowed to provide arguments as to why some indicators are more relevant than others in their situation (flexibility). Participants were also told that

they could indicate if the measurement of some performance indicators should be changed (repair), and that they would receive feedback on the overall teaching quality in their school (internal transparency). Overall, this experimental condition captures the enabling characteristics of a performance measurement system (Adler & Borys, 1996).

Participants allocated to the *Coercive Design Characteristics* condition learned that the purpose of the system is to make them comply with the school's standards for good teaching. The performance goals are not clearly defined, and the procedure does not allow them to provide arguments as to why some indicators are more relevant than others in their situation. Participants were further told that they could not indicate that the measurement of some performance indicators should be changed, and they would not receive any feedback on the overall teaching quality in their school. As such, the coercive nature of the PMS is captured by a top-down focus that does not allow employee input or flexibility when working with the system (Ahrens & Chapman, 2004; Hoy & Sweetland, 2001).

To manipulate the development process, we presented information to participants about how the PMS was developed. In the *Enabling Development* condition, participants learned that they were very much involved, that they were given many opportunities to provide input, and that they were given the possibility of testing, reviewing, and refining the new system before it became operational. In the *Coercive Development* condition, participants read that they were not at all involved, that they were not given any opportunities to provide input, and that they had no possibility to test, review, or refine the new system before it became operational. The texts of these vignettes are provided in "Appendix".

The survey experiment was conducted using two distinct samples. Study 1 is a laboratory study using a (mostly Western European) student sample. This sample allowed us to achieve good internal validity through strong control of the treatments the participants were subjected to. The use of students as experimental subjects is much debated in the research community. Notably, concerns have been raised about how well students represent the general population (e.g. Falk & Heckman, 2009; Thomassen et al., 2017), although there is a lack of evidence that there are marked differences between students and, for instance, practitioners in public management (Walker et al., 2017). To address this concern, we have followed the suggestion of others and adopted empirical replication as this enables a systematic comparison of findings across different samples (Walker et al., 2017). We therefore replicated our experiment in Study 2 using the online crowdsourcing platform MTurk with US participants. MTurk enables scholars to collect high-quality data in a cost-effective way (Buhrmester et al., 2011; Germine et al., 2012). Scholars opt to use this platform as it provides a better sample quality than student pools, offers access to a pool of workers with diverse backgrounds, and can thus broaden the validity of a study beyond the student population (Mason & Suri, 2012). The second experiment is thus a replication of the first experiment but with a different sample. This type of replication not only allows us to address potential validity issues with an isolated student or MTurk sample, but also to examine to what extent findings are generalizable to different countries and participants with different backgrounds.

The power analysis program G*power (Faul et al., 2009) was used to predetermine the required sample sizes. We determined that, with a medium effect size of $d=0.25$ and power of 80%, a sample size of 128 would be required.

3.2 Variables

For our first main dependent variable, *procedural fairness*, we used three items adapted from Van Prooijen et al. (2002). These items asked “how fair/just/appropriate was the performance measurement system used in your school for evaluating your performance” (1 = absolutely not; 7 = absolutely). These questions were combined into a procedural fairness scale.

For our second main dependent variable, *red tape* (as related to the performance measurement system), we used the three-item red tape (TIRT) scale suggested by Borry (2016). This scale asked participants to rate the performance measurement procedure on three dimensions, namely effectiveness, necessity, and burden, using seven-point Likert scales (where 1 = very ineffective/unnecessary/not at all burdensome and 7 = very effective/necessary/very burdensome).

Further, to check that our manipulations had been effective, we asked two questions: “Were you involved in the development of the performance measurement system used in your school?” (Yes/No) and “What was the stated purpose of the performance measurement system used in your school?” (To make you comply with the school’s standards for good teaching/to help you become a better lecturer). Furthermore, we used an adapted version of the Instructional Manipulation Check (Oppenheimer et al., 2009) to check whether people were carefully following instructions. Finally, we asked participants for some background characteristics, including their age, gender, and working experience. See “Appendix” for a full overview of the measurements used in this study.

4 Results

4.1 Study 1: Student sample

4.1.1 Participants

We conducted the experiment in a research lab at a Dutch research university, with a total of 208 student participants. The online experiment was created in Qualtrics. Participants in this between-subjects factorial design were randomly assigned to either the Enabling or the Coercive Design Characteristics condition, and to the Enabling or the Coercive Development condition. This study formed the first part of a series of studies that in total took one hour to complete. Our specific study took between 10 and 15 min to complete. As an incentive to participate in the one-hour series of studies, students were either rewarded with 8 euros or given course credits for participation (depending on their choice). Participants completed the online survey in one-person cubicles. Six participants were removed from the dataset because

Table 1 Background characteristics of student sample

	N	% Male	Age	Working experience
Group 1 (Coercive DC; Coercive DP)	52	52%	21.23	4.08
Group 2 (Coercive DC; Enabling DP)	31	38%	22.45	4.72
Group 3 (Enabling DC; Coercive DP)	33	55%	21.64	4.55
Group 4 (Enabling DC; Enabling DP)	48	44%	21.56	4.13
F		0.76	1.38	0.45
Prob. > F		0.52	0.25	0.72

DC design characteristics, *DP* development process

they failed to fully complete the surveys. A further 28 participants failed the attention check question and were subsequently removed from the sample. An additional 10 participants failed one or both of the manipulation checks and were therefore also excluded. This meant that our final sample consisted of 164 participants.¹

We checked for randomization based on age, gender, and working experience across the four treatments.² As shown in Table 1, randomization on these background characteristics was successful.³ The average age in our sample was 21.66 years ($SD=2.89$), and 52.4% of our participants were male. Participants had little working experience: on average 4.3 years ($SD=2.89$). This is to be expected for a Dutch student sample.

4.1.2 Results

We first calculated the procedural fairness scores by taking the average of the three fairness item scores. An Exploratory Factor Analysis yielded one factor (Eigenvalue=2.42; 80.50% explained variance) and the scale's reliability was high ($\alpha=0.92$). We then calculated the TIRT scores by averaging the three items on rule effectiveness (reversed coded), necessity (reversed coded), and burden. Although the Exploratory Factor Analysis of the three items again yielded a single factor (Eigenvalue=1.28; 42.64% explained variance), the scale had a relatively poor reliability ($\alpha=0.62$).⁴ A summary of all the cell means and standard deviations is provided in Table 2 (panel A).⁵ Results of the ANOVA analysis are also shown in Table 2 (panel B).

¹ We re-ran the analyses without excluding those participants who failed the manipulation check and/or the attention check, and results are similar to those reported in the paper.

² Two participants did not provide demographic information but did complete the rest of the experiment. These participants are included in the final sample.

³ Note that due to failures on the attention check, cell sizes were not equal. Unequal cell sizes can mean that the F test is not robust against violations of normality. However, we also combined the two datasets, resulting in more equal cell sizes, and found similar effects. See also the discussion at the end of the results of Study 2.

⁴ For this reason, we also created a two-item red tape scale (combining rule effectiveness and necessity that had a higher reliability, $\alpha=0.73$). Nevertheless, the results for this two-item red tape scale are very similar to those for the full TIRT scale which is reported here.

⁵ The red tape and procedural fairness scales were significantly negatively correlated, $r=-.71$. However, since, theoretically, the two concepts are distinct, we present separate analyses for these two scales.

Table 2 Results: student sample*Panel A: Summary of means (standard deviations) for the dependent variables*

Dependent variable	Design characteristics	Total	Development process	
			Coercive	Enabling
Procedural fairness	Coercive	3.22 (1.38)	2.82 (1.26)	3.89 (1.34)
	Enabling	5.27 (1.02)	4.90 (1.05)	5.52 (.92)
	Total	4.23 (1.59)	3.63 (1.56)	4.99 (1.36)
Red tape	Coercive	4.38 (1.07)	4.63 (1.07)	3.96 (1.11)
	Enabling	3.22 (.82)	3.60 (.92)	2.97 (.63)
	Total	3.81 (1.11)	4.23 (1.07)	3.35 (.97)

Panel B: ANOVA results: experiment 1

Dependent variable	Main/interaction effect
Procedural fairness	
Design characteristics	$F(1,160)=102.10, p<.001$
Development process	$F(1,160)=21.32, p<.001$
Interaction effect	$F(1,160)=1.51, p=.222$
Red tape	
Design characteristics	$F(1,160)=49.27, p<.001$
Development process	$F(1,160)=20.46, p<.001$
Interaction effect	$F(1,160)=0.03, p=.871$

An ANOVA on the procedural fairness scale yielded a significant main effect of Design Characteristics ($F(1,160)=102.10, p<0.001$, partial $\eta^2=39$). That is, a performance measurement system with enabling design characteristics resulted in higher perceived procedural fairness ($M=5.27, SD=1.02$) than a system with coercive design characteristics ($M=3.22, SD=1.38$). We also found a significant main effect of Development Process ($F(1,160)=21.32, p<0.001$, partial $\eta^2=0.12$). Participants in the enabling development condition perceived higher procedural fairness ($M=4.88, SD=1.36$) compared to participants in the coercive development condition ($M=3.63, SD=1.56$). There was no significant interaction effect for Development Process and Design Characteristics ($F(1,160)=1.51, p=0.222$, partial $\eta^2=0.01$).

An ANOVA on the TIRT scale yielded a significant main effect of Design Characteristics ($F(1,160)=49.27, p<0.001$, partial $\eta^2=0.24$). A performance measurement system with enabling design characteristics was associated with less perceived red tape ($M=3.22, SD=0.82$) than a system with coercive design characteristics ($M=4.38, SD=1.07$). We also found a significant main effect of Development Process ($F(1,160)=20.46, p<0.001$, partial $\eta^2=0.11$). Those individuals who were

Table 3 Background characteristics of MTurk sample

	N	% Male	Age	Working experience
Group 1 (Coercive DC; Coercive DP)	43	51%	42.47	20.93
Group 2 (Coercive DC; Enabling DP)	50	52%	38.42	18.54
Group 3 (Enabling DC; Coercive DP)	55	51%	40.29	19.91
Group 4 (Enabling DC; Enabling DP)	41	45%	40.20	17.85
F		0.24	0.86	0.60
Prob. > F		0.87	0.47	0.62

DC design characteristics, *DP* development process

told they had been involved in the development process of the performance measurement system (an enabling development process) perceived less red tape ($M=3.35$, $SD=0.97$) than those who were told they were not involved (a coercive development process) ($M=4.23$, $SD=1.07$). There was no significant interaction effect for Development Process and Design Characteristics ($F(1,160)=0.03$, $p=0.871$, partial $\eta^2=0.00$).

Overall, our findings provide strong support for our hypotheses that a performance measurement system with enabling design characteristics leads to perceptions of greater procedural fairness (Hypothesis 1a) and less red tape (Hypothesis 2a) than a system with coercive design characteristics. We also found support for Hypotheses 1b and 2b, which state that an enabling development process leads to perceptions of greater procedural fairness and less red tape than a coercive development process.

4.2 Study 2: Online citizen sample (MTurk)

4.2.1 Participants

Study 2 adopted the same design as used in Study 1 but, this time, the experiment was administered using a citizen sample from Amazon's MTurk rather than a student sample. In total, 203 individuals participated. Participation in the experiment was limited to participants located in the United States. Participants were required to have a 99% approval rating on previous MTurk assignments, and the study took about 10 min to complete. MTurkers were paid \$1.50 for completing the study. This payment rate is lower than in the student sample but somewhat higher than that usually seen at that time on MTurk. However, as expected given their reputation concerns (Peer et al., 2014), and in line with prior studies (Hauser & Schwarz, 2016), participants from the MTurk sample appeared to pay closer attention to the provided instructions than our student sample. Two participants did not complete the survey and were therefore excluded from the sample. One participant failed the attention check question and was also excluded. Finally, 11 further participants failed one or

Table 4 Results: MTurk sample*Panel A: Summary of means (standard deviations) for the dependent variables*

Dependent variable	Design characteristics	Total	Development Process	
			Coercive	Enabling
Procedural fairness	Coercive	3.03 (1.44)	2.32 (.95)	3.64 (1.51)
	Enabling	5.04 (1.54)	4.67 (1.55)	5.54 (1.41)
	Total	4.05 (1.80)	3.64 (1.76)	4.49 (1.74)
Red tape	Coercive	4.85 (1.25)	5.24 (.88)	4.51 (1.42)
	Enabling	3.79 (1.18)	3.99 (1.25)	3.52 (1.03)
	Total	4.32 (1.32)	4.54 (1.26)	4.07 (1.35)

Panel B: ANOVA results experiment 2

Dependent variable	Main/interaction effect
Procedural fairness	
Design characteristics	$F(1,185) = 108.82, p < .001$
Development process	$F(1,185) = 28.77, p < .001$
Interaction effect	$F(1,185) = 1.27, p = .262$
Red tape	
Design characteristics	$F(1,185) = 41.82, p < .001$
Development process	$F(1,185) = 12.02, p < .01$
Interaction effect	$F(1,185) = .54, p = .465$

both manipulation checks and were therefore excluded. As a result, our final dataset consists of 189 participants.⁶

We checked that we had adequately randomized the participants in terms of age, gender, and working experience across the four treatments.⁷ As shown in Table 3, the randomization based on these background characteristics was successful. The average age in our sample was 40.24 years ($SD = 12.16$), and 50.3% of our participants were male. Participants had, on average, 19.34 years ($SD = 11.68$) of working experience.

4.2.2 Results

The procedural fairness scale for the MTurk sample had a high reliability ($\alpha = 0.97$). As in Study 1, the EFA yielded a single factor (Eigenvalue = 2.75; 91.57% explained variance). Similarly, an EFA on the three items of the TIRT scale for the MTurk

⁶ We reran the analyses including those participants who failed the manipulation check and/or the attention check, and the results were similar to those reported in the paper.

⁷ One participant did not provide information on working experience, but did complete the rest of the experiment, and was retained in the final sample.

sample yielded one factor (Eigenvalue = 1.50; 50.03% explained variance) and the scale had an acceptable reliability ($\alpha = 0.74$).⁸ A summary of all the cell means and standard deviations is provided in Table 4 (panel A). Results of the ANOVA analysis are also shown in Table 4 (panel B).

An ANOVA on the procedural fairness scale showed that Design Characteristics had a significant main effect ($F(1,185) = 108.82, p < 0.001$, partial $\eta^2 = 0.37$). That is, a performance measurement system with enabling design characteristics resulted in higher perceived procedural fairness ($M = 5.04, SD = 1.54$) than a system with coercive design characteristics ($M = 3.03, SD = 1.44$). We also found a significant main effect of Development Process ($F(1,185) = 28.77, p < 0.001$, partial $\eta^2 = 0.14$). Participants in the enabling development condition perceived higher procedural fairness ($M = 4.49, SD = 1.74$) than those in the coercive development condition ($M = 3.64, SD = 1.76$). There was no significant interaction effect for Development Process and Design Characteristics ($F(1,185) = 1.27, p = 0.262$, partial $\eta^2 = 0.01$).

An ANOVA on the TIRT scale yielded a significant main effect of Design Characteristics ($F(1,185) = 41.82, p < 0.001$, partial $\eta^2 = 0.18$). That is, a performance measurement system with enabling design characteristics was perceived to entail less red tape ($M = 3.79, SD = 1.18$) than a system with coercive design characteristics ($M = 4.85, SD = 1.25$). We also found a significant main effect of Development Process ($F(1,185) = 12.02, p < 0.01$, partial $\eta^2 = 0.06$). Those individuals who were told that they were involved in the development process of the performance measurement system (enabling development) perceived the system to entail less red tape ($M = 4.07, SD = 1.35$) than those who were told they were not involved (coercive development) ($M = 4.54, SD = 1.26$). There was no significant interaction effect for Development Process and Design Characteristics ($F(1,185) = 0.54, p = 0.465$, partial $\eta^2 = 0.00$).

These findings support Hypotheses 1a and 2a, which state that a performance measurement system with enabling design characteristics leads to perceptions of greater procedural fairness and less red tape than a system with coercive design characteristics. We also found support for Hypotheses 1b and 2b, which state that an enabling development process leads to a higher level of perceived procedural fairness and a lower level of perceived red tape than a coercive development process.

Overall, the main findings from the two samples are quite similar. In line with our expectations, we consistently found that enabling design characteristics as well as enabling development processes lead to perceptions of greater procedural fairness and less red tape than coercive design characteristics and a coercive development process. Note that we also combined the two datasets and ran ANCOVAs with Sample (students vs. MTurk) as covariate. These analyses yielded similar effects as described above. In addition, the ANCOVA on perceived red tape showed a significant effect of the covariate, $F(1,348) = 21.55, p < 0.001$. However, a $2 \times 2 \times 2$ ANOVA including Sample as an additional, third factor only yielded a main effect of Sample ($M_{\text{student}} = 3.81$ vs. $M_{\text{MTurk}} = 4.31$), with Sample not interacting with either

⁸ As in Study 1, red tape and procedural fairness were significantly negatively correlated, $r = -.78$. As the two concepts are theoretically different, we present separate analyses for each measure.

Development Process or Design Characteristics. As such, there is support for the claim that our findings can be generalized across samples.

5 Discussion and conclusions

This study adds to an emerging stream of literature on enabling versus coercive systems while taking the perspective of the employees who are subject to such a system (Bisbe et al., 2019) and investigates the consequences of enabling versus coercive systems on employees' perceptions of the procedural quality of these systems. Drawing on research from the management control and organization studies literature streams, we take the first step in understanding the effects of enabling versus coercive performance measurement system design characteristics and of enabling versus coercive development processes on perceptions of procedural quality, captured by procedural fairness and red tape. In a survey experiment using a laboratory student sample and an online sample using Amazon's MTurk, these employee perceptions are found to be more positive with enabling systems than with coercive systems, which supports our hypotheses and the underlying reasoning.

In line with our hypotheses, we consistently found that a performance measurement system that is designed with enabling characteristics (in terms of repair, flexibility, internal transparency, and global transparency) is perceived as having higher procedural fairness and less red tape than a system with coercive characteristics (and therefore lacking these qualities). We also found that a performance measurement system that has an enabling development process, that involves employees' participation, leads to higher procedural fairness and less red tape than a coercive development process that excludes employee involvement.

Further, we found that these two distinct aspects did not affect the strength of each other's effect. In other words, design characteristics did not interact with the development process: both effects were independent. In addressing this, the study examined the joint effects of the design characteristics and development process in an exploratory way because the theoretical arguments were not clear cut. On the one hand, an enabling development process might reduce the negative effects of a coercive design since employees may internalize the value of the performance measurement system (Deci & Ryan, 2000). On the other hand, an enabling development process may result in a better appreciation of an enabling design, particularly if the design reflects a good understanding of the organizational setting and the problems at hand (Wouters & Wilderom, 2008). The lack of significant interaction effects could indicate that the suggested effects do not exist or, alternatively, that they cancel each other out. It is also important to note that the experimental design did not allow changes in the performance measurement design based on the development process. In a real-life setting, it is more likely that the design of the performance measurement system will reflect the knowledge and understanding of the people involved in the development process. Further, the internalization of the value of the system may also be stronger. Future research could study these potential processes in more detail.

Our study contributes to the performance measurement and management control literature in several ways. First, as suggested by Tessier and Otley (2012), we take a tentative step towards opening up the black box of enabling or coercive types of performance measurement by distinguishing between enabling versus coercive performance measurement on the one hand and the perceived procedural quality on the other. As these authors argue, recognizing this distinction is important since it is related to the difference between managerial intentions—as reflected in an enabling versus coercive system—and employee perceptions of the system. By incorporating this distinction in our study, we are able to empirically examine how enabling and coercive performance measurement systems affect procedural quality as captured by procedural fairness and perceived red tape, and to increase our understanding of how individuals respond to performance measurement systems by examining individual responses to such systems (see also Van der Kolk & Kaufmann, 2018; Van der Kolk et al., 2019). Rather than derive enabling and coercive control perceptions from qualitative case studies (e.g. Ahrens & Chapman, 2004; Wouters & Wilderom, 2008), we have focused on a single procedure that is manipulated experimentally to study enabling and coercive performance measurement systems. This approach allows us to tease out causal mechanisms and make claims about specific relationships between enabling and coercive design characteristics and the development processes of the performance measurement system on the one hand, and indicators of perceived procedural quality such as procedural fairness and red tape on the other. Given that a single sample would have been sufficient to test the hypotheses, the empirical replication (Walker et al., 2017) of the experiment can be seen as a strong additional feature of this study as it is helpful in addressing possible validity concerns related to the experiment. While the experiment was designed for the purposes of this specific study, it may well be used in future research on the effects of enabling versus coercive systems. The replication is also important for addressing concerns related to how well the participants represent the population and/or low- or mid-level employees.

Second, our findings hint at different options for improving procedural quality related to performance measurement. Our findings show that the perceptions of procedural quality are influenced both by design characteristics of the performance measurement system and by the type of process used to develop the system. As such, our research adds to studies that focus only on design characteristics (e.g. Ahrens & Chapman, 2004) or on the development process (e.g. Wouters & Wilderom, 2008). In addition, the insights into the effect of enabling versus coercive systems on procedural quality are novel and add to the existing knowledge because the orientation of the system, and in particular of the design characteristics, extends beyond a high procedural quality. Managers and policymakers interested in improving procedural quality related to performance measurement systems could choose to amend existing systems to include more enabling characteristics, or to stimulate employee participation in the development of new or adjusted systems. Our results clearly show that both approaches contribute to this end, although in our experiment addressing design characteristics appears to be the more effective of the two. If it is not feasible to involve employees in the development of the performance measurement system,

including more enabling characteristics in the performance measurement system could serve as an effective alternative.

This study has a number of limitations that could be addressed in future research. First, the use of an experimental design is always associated with concerns about external validity. Indeed, we cannot be certain that the findings from our experiments are representative of employees in general, or business school lecturers' attitudes in particular. We did, however, empirically replicate the study using two different samples to test our hypotheses (a student sample and a citizen sample) with similar procedures. The two samples provided similar results, suggesting that at least to some extent the results are generalizable. Future research could use this or a similar experimental design to further investigate the relationship between enabling and coercive types of performance measurement and other relevant outcomes.

Second, while this research constitutes a first step toward increasing the conceptual clarity of an enabling versus coercive orientation of a performance measurement system, by separating this orientation from an assessment of the procedural quality of the system, the experiment did not empirically establish a difference between orientation and procedural quality. This would have been the case when an enabling system is perceived negatively or when a coercive system is perceived positively. An interesting avenue for future research would therefore be to extend this study by including other factors that may influence how the performance measurement system is perceived, such as trust or communication of the managerial intention when designing the system (see, for example, Väisänen et al., 2020) or the presentation format used to report the performance measures (see, for example, Cardinaels & Van Veen-Dirks, 2010) and to see how these interact with an enabling or coercive orientation of the performance measurement system. Furthermore, beyond having effects on employees' perceptions and attitudes, performance measurement systems fulfill purposes, in particular decision facilitating and decision influencing purposes (Sprinkle, 2003; Van Veen-Dirks, 2010). Future research could address the question how enabling versus coercive design characteristics or development processes influence the possibilities for a performance measurement system to achieve one or both purposes. Studying how both purposes can be fulfilled seems relevant also for bridging the gap between research streams that address either one or the other purpose (see also Luft, 2016).

Third, we have created a fictitious performance measurement procedure with no overlap between design characteristics and the development process to tease out the independent effects on employees' attitudes to the system. Such a stylized approach is very common in experimental research but may not well reflect a real-life setting. In practice, managers are likely to opt for different combinations of enabling and coercive design characteristics and development processes (due to financial or time constraints for example). Future studies could consider using identified real-life procedures as a starting point for analyzing the effects of performance measurement systems on employees whose performance is being measured. In such a study, the participants could be people actually working with the systems being investigated in day-to-day practice, and any influence they have on the design characteristics during the development process of a performance measurement procedure could be

investigated in detail. The use of real-life procedures would also address some of the general concerns about the generalizability of experimental studies.

Appendix: Text of vignettes and measurements

Introduction (identical across treatments)

You have been working as a lecturer at a large business school. The school has recently introduced a new performance measurement system, which is used to monitor the school's teaching quality. The new system is used throughout the school for all lecturers.

The performance measurement procedure requires you to fill out and submit four times a year a six-page document which contains an overview of all your activities. In this "Teaching Review Document", you have to provide an overview of, and reflect upon, your teaching activities. Completing this document will take about four hours each submission.

Manipulation 1: Development of performance measurement system

Enabling

You were very much involved in the development of the new performance measurement system from the very beginning of the process.

You were given many opportunities to provide input to the new system based on your own working experience during the initial design stage that started 3 years ago.

The school had also given you the possibility of testing, reviewing, and refining the new system before it became operational last year.

Coercive

You were not at all involved at any point in the development of the new performance measurement system.

You were not given any opportunity to provide input to the new system based on your own working experience during the initial design stage that started 3 years ago.

The school had also not given you the possibility of testing, reviewing, or refining the new system before it became operational last year.

Manipulation 2: Characteristics of performance measurement system

Enabling

You are evaluated on several dimensions of teaching quality.

There are four performance indicators, namely:

- (1) evaluations from students
- (2) evaluations from colleagues (also known as “peer review”)
- (3) teaching certificates and training
- (4) focus groups with student representatives

The purpose of the system is to help you become a better lecturer.

The performance goals for each indicator are clearly defined, which helps you to know what is expected of you in your position. The indicators allow for a transparent assessment of how well you have performed.

You can provide arguments in the “Teaching Review Document” as to why some indicators are more relevant than others in your situation.

You are also able to indicate in the “Teaching Review Document” if the measurement of some performance indicators should be changed in your case.

You will receive a short document with feedback on the overall teaching quality in your school four times a year.

Coercive

You are evaluated on several dimensions of teaching quality.

There are four performance indicators, namely:

- (1) evaluations from students
- (2) evaluations from colleagues (also known as “peer review”)
- (3) teaching certificates and training
- (4) focus groups with student representatives

The purpose of the system is to make you comply with the school’s standards for good teaching.

The performance goals for the indicators are not clearly defined, which makes it difficult to know what is expected of you in your position. This means that the assessment of how well you have performed is not transparent.

All the indicators are considered to be equally important. You are unable to provide arguments in the “Teaching Review Document” as to why some indicators are more relevant than others in your situation.

The school also decides on how the performance indicators are measured. There is no possibility for you to indicate in the “Teaching Review Document” if the measurement of some performance indicators should be changed in your situation.

You do not receive any feedback on the overall teaching quality in your school

Measurements

Introduction to questions	<p>Questions about the performance measurement system</p> <p>You will now be asked a number of questions about the performance measurement system used in your school. Keep in mind that we ask you to assume the role of a lecturer at a large business school when answering these questions. It is important that you choose the answers that you feel are most appropriate given your role as lecturer.</p>												
Red Tape (TIRT scale)	<p>In your view, how effective is the performance measurement system used in your school? (1 = very ineffective; 7 = very effective)</p> <p>In your view, how necessary is the performance measurement system used in your school? (1 = very unnecessary; 7 = very necessary)</p> <p>In your view, how burdensome is the performance measurement system used in your school? (1 = not at all burdensome; 7 = very burdensome)</p>												
Procedural fairness	<p>How fair is the performance measurement system used in your school for evaluating your performance? (1 = unfair; 7 = fair)</p> <p>How just is the performance measurement system used in your school for evaluating your performance? (1 = unjust; 7 = just)</p> <p>How appropriate is the performance measurement system used in your school for evaluating your performance? (1 = inappropriate; 7 = appropriate)</p>												
Manipulation checks	<p>Were you involved in the development of the performance measurement system used in your school? Yes/No</p> <p>What was the stated purpose of the performance measurement system used in your school? 1 = to help you become a better lecturer, 2 = to make you comply with the school's standards for good teaching</p>												
Demographics	<p>What is your gender, age, in which country were you born, what is your native language, in general how would you describe your political views (1—very liberal; 7 = very conservative), how many years of working experience do you have?</p>												
Attention check	<p>Organizational culture is a fuzzy concept that is hard to define. To help us understand how people interact in organizations we are interested in how people react to culture. Specifically, we are interested in how well you read instructions; if not, your answers may not tell us much about people in real organizations. To show that you have read these instructions please ignore the question below about organizational culture and check only "None of the above" as your answer</p> <p>Please select all that describe the organizational culture that fits your personality best:</p> <table border="0" style="width: 100%;"> <tr> <td><input type="checkbox"/> Fun</td> <td><input type="checkbox"/> Free</td> </tr> <tr> <td><input type="checkbox"/> Exciting</td> <td><input type="checkbox"/> Tolerant</td> </tr> <tr> <td><input type="checkbox"/> Dreadful</td> <td><input type="checkbox"/> Obedient</td> </tr> <tr> <td><input type="checkbox"/> Innovative</td> <td><input type="checkbox"/> Oppressive</td> </tr> <tr> <td><input type="checkbox"/> Collaborative</td> <td><input type="checkbox"/> None of the above</td> </tr> <tr> <td><input type="checkbox"/> Open</td> <td></td> </tr> </table>	<input type="checkbox"/> Fun	<input type="checkbox"/> Free	<input type="checkbox"/> Exciting	<input type="checkbox"/> Tolerant	<input type="checkbox"/> Dreadful	<input type="checkbox"/> Obedient	<input type="checkbox"/> Innovative	<input type="checkbox"/> Oppressive	<input type="checkbox"/> Collaborative	<input type="checkbox"/> None of the above	<input type="checkbox"/> Open	
<input type="checkbox"/> Fun	<input type="checkbox"/> Free												
<input type="checkbox"/> Exciting	<input type="checkbox"/> Tolerant												
<input type="checkbox"/> Dreadful	<input type="checkbox"/> Obedient												
<input type="checkbox"/> Innovative	<input type="checkbox"/> Oppressive												
<input type="checkbox"/> Collaborative	<input type="checkbox"/> None of the above												
<input type="checkbox"/> Open													

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adler, P. S. (1999). Building better bureaucracies. *Academy of Management Executive*, 13(4), 36–47.
- Adler, P. S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly*, 41(1), 61–89.
- Ahrens, T., & Chapman, C. S. (2004). Accounting for flexibility and efficiency: A field study of management control systems in a restaurant chain. *Contemporary Accounting Research*, 21(2), 271–301.
- Bellavance, F., Landry, S., & Schiehl, E. (2013). Procedural justice in managerial performance evaluation: Effects of subjectivity, relationship quality, and voice opportunity. *The British Accounting Review*, 45(3), 149–166.
- Bisbe, J., Kruis, A.-M., & Madini, P. (2019). Coercive, enabling, diagnostic, and interactive control: Untangling the threads of their connections. *Journal of Accounting Literature*, 43, 124–144.
- Borror, E. L. (2016). A new measure of red tape: Introducing the three-item red tape (TIRT) scale. *International Public Management Journal*, 19(4), 573–593.
- Bozeman, B. (1993). A theory of government “red tape.” *Journal of Public Administration Research and Theory*, 3(3), 273–304.
- Bozeman, B. (2000). *Bureaucracy and red tape*. Upper Saddle River, NJ: Prentice Hall.
- Bozeman, B., & Anderson, D. M. (2016). Public policy and the origins of bureaucratic red tape: Implications of the Stanford yacht scandal. *Administration & Society*, 48(6), 736–759.
- Bozeman, B., & Feeney, M. K. (2011). *Rules and red tape: A prism for public administration theory and research*. Armonk, NY: M.E. Sharpe.
- Brewer, G. A., & Walker, R. M. (2010). The impact of red tape on governmental performance: An empirical analysis. *Journal of Public Administration Research and Theory*, 20(1), 233–257.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3–5.
- Cardinaels, E., & Van Veen-Dirks, P. (2010). Financial versus non-financial information: The impact of information organization and presentation in a Balanced Scorecard. *Accounting, Organizations and Society*, 35(6), 565–578.
- Chapman, C. S., & Kihn, L. A. (2009). Information system integration, enabling control and performance. *Accounting, Organizations and Society*, 34(2), 151–169.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86(3), 425–445.
- Colquitt, J. A., & Jackson, C. L. (2006). Justice in teams: The context sensitivity of justice rules across individual and team contexts 1. *Journal of Applied Social Psychology*, 36(4), 868–899.
- Cugueró-Escofet, N., & Rosanas, J. M. (2013). The just design and use of management control systems as requirements for goal congruence. *Management Accounting Research*, 24(1), 23–40.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.
- DeHart-Davis, L. (2009). Green tape: A theory of effective organizational rules. *Journal of Public Administration Research and Theory*, 19(2), 361–384.
- DeHart-Davis, L., & Pandey, S. K. (2005). Red tape and public employees: Does perceived rule dysfunction alienate managers? *Journal of Public Administration Research and Theory*, 15(1), 133–148.
- Dolan, P., Edlin, R., Tsuchiya, A., & Wailoo, A. (2007). It ain't what you do, it's the way that you do it: Characteristics of procedural justice and their importance in social decision-making. *Journal of Economic Behavior & Organization*, 64(1), 157–170.

- Dowling, C., & Leech, S. A. (2014). A Big 4 firm's use of information technology to control the audit process: How an audit support system is changing auditor behavior. *Contemporary Accounting Research*, *31*(1), 230–252.
- Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, *326*(5952), 535–538.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.
- Folger, R. G., & Cropanzano, R. (1998). *Organizational justice and human resource management* (Vol. 7). Thousand Oaks: Sage.
- Germiné, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857.
- Groen, B. A. (2018). A survey study into participation in goal setting, fairness, and goal commitment: Effects of including multiple types of fairness. *Journal of Management Accounting Research*, *30*(2), 207–240.
- Groen, B. A., Wouters, M. J., & Wilderom, C. P. (2012). Why do employees take more initiatives to improve their performance after co-developing performance measures? A field study. *Management Accounting Research*, *23*(2), 120–141.
- Hartmann, F., & Slapničar, S. (2012). The perceived fairness of performance evaluation: The role of uncertainty. *Management Accounting Research*, *23*(1), 17–33.
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.
- Hoy, W. K., & Sweetland, S. R. (2001). Designing better schools: The meaning and measure of enabling school structures. *Educational Administration Quarterly*, *37*(3), 296–321.
- Jordan, S., & Messner, M. (2012). Enabling control and the problem of incomplete performance indicators. *Accounting, Organizations and Society*, *37*(8), 544–564.
- Jorgenson, B., & Messner, M. (2009). Management control in new product development: The dynamics of managing flexibility and efficiency. *Journal of Management Accounting Research*, *21*(1), 99–124.
- Kaufmann, W., & Feeny, M. K. (2014). Beyond the rules: The effect of outcome favourability on red tape perceptions. *Public Administration*, *92*(1), 178–191.
- Kaufmann, W., & Tummers, L. G. (2017). The negative effect of red tape on procedural satisfaction. *Public Management Review*, *19*(9), 1311–1327.
- Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27–55). New York, NY: Springer.
- Leventhal, G. S., Karuza, J., & Fry, W. R. (1980). Beyond fairness: A theory of allocation preferences. *Justice and Social Interaction*, *3*(1), 167–218.
- Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York: Plenum Press.
- Luft, J. (2016). Management accounting in the laboratory and in social context. Four contrasts, 1975–2014. *Accounting, Organizations and Society*, *49*, 9–20.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.
- Mundy, J. (2010). Creating dynamic tensions through a balanced use of management control systems. *Accounting, Organizations and Society*, *35*(5), 499–523.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031.
- Pfister, J. A., & Lukka, K. (2019). Interrelation of controls for autonomous motivation: A field study of productivity gains through pressure-induced process innovation. *The Accounting Review*, *94*(3), 345–371.
- Proenca, T. (2010). Self-managed work teams: An enabling or coercive nature. *The International Journal of Human Resource Management*, *21*(3), 337–354.
- Sprinkle, G. B. (2003). Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, *28*(2–3), 287–318.
- Tessier, S., & Otley, D. (2012). A conceptual development of Simons' Levers of Control framework. *Management Accounting Research*, *23*(3), 171–185.

- Thomassen, J. P., Leliveld, M. C., Van de Walle, S., & Ahaus, K. (2017). Compensating citizens for poor service delivery: Experimental research in public and private settings. *Public Administration*, 95(4), 895–911.
- Väisänen, M., Tessier, S., & Järvinen, J. (2020). Fostering enabling perceptions of management controls during post-acquisition integration. *Contemporary Accounting Research* (Forthcoming).
- Van Prooijen, J. W., Van den Bos, K., & Wilke, H. A. M. (2002). Procedural justice and status: Status salience as antecedent of procedural fairness effects. *Journal of Personality and Social Psychology*, 83, 1353–1361.
- Van der Kolk, B., & Kaufmann, W. (2018). Performance measurement, cognitive dissonance and coping strategies: Exploring individual responses to NPM-inspired output control. *Journal of Management Control*, 29, 93–113.
- Van der Kolk, B., Van Veen-Dirks, P., & Ter Bogt, H. J. (2019). The impact of management control on employee motivation and performance in the public sector. *European Accounting Review*, 28(5), 901–928.
- Van Veen-Dirks, P. (2010). Different uses of performance measures: The evaluation versus reward of production managers. *Accounting, Organizations and Society*, 35(2), 141–164.
- Voußem, L., Kramer, S., & Schäffer, U. (2016). Fairness perceptions of annual bonus payments: The effects of subjective performance measures and the achievement of bonus targets. *Management Accounting Research*, 30, 32–46.
- Walker, R. M., James, O., & Brewer, G. A. (2017). Replication, experiments and knowledge in public management research. *Public Management Review*, 19(9), 1221–1234.
- Wong-On-Wing, B., Guo, L., & Lui, G. (2010). Intrinsic and extrinsic motivation and participation in budgeting: Antecedents and consequences. *Behavioral Research in Accounting*, 22(2), 133–153.
- Wouters, M. (2009). A developmental approach to performance measures—Results from a longitudinal case study. *European Management Journal*, 27(1), 64–78.
- Wouters, M., & Roijmans, D. (2011). Using prototypes to induce experimentation and knowledge integration in the development of enabling accounting information. *Contemporary Accounting Research*, 28(2), 708–736.
- Wouters, M., & Wilderom, C. (2008). Developing performance-measurement systems as enabling formalization: A longitudinal field study of a logistics department. *Accounting, Organizations and Society*, 33(4–5), 488–516.