Management Accounting Research 23 (2012) 120–141

Contents lists available at SciVerse ScienceDirect

Management Accounting Research

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

Why do employees take more initiatives to improve their performance after co-developing performance measures? A field study

Bianca A.C. Groen*, Marc J.F. Wouters[1], Celeste P.M. Wilderom

University of Twente, School of Management and Governance, P.O. Box 217, Enschede, Netherlands

ARTICLE INFO

Keywords:
Employee participation
Performance measurement
Theory of planned behavior
Operations management

ABSTRACT

Performance measurements may stimulate employee initiatives to improve operational performance, especially when employees themselves participate in the development of their own departmental performance measures. Using the theory of planned behavior, we examine why this occurs in a beverage manufacturing company where we helped bottling line maintenance technicians develop measures about the results of their own work. Our analyses are based on qualitative data gathered at 156 meetings, 34 semi-structured interviews, quantitative performance data from the company's information systems, and quantitative questionnaire data. We found that the participatory development process increased employees' attitude, perceived social pressure and perceived capability to take initiative. Moreover, the departmental performance improved when the jointly developed performance measures were put to use.

1. Introduction

The participation of employees is an important theme in management accounting research (e.g., Derfuss, 2009; Luft and Shields, 2007). Most studies investigate participation in budgeting: the amount of influence a subordinate manager has for setting his/her unit's budgets. Derfuss (2009) conducted a meta-analysis and found 11 positive consequences of participative budgeting that generalize across samples (e.g., the positive effect of budgetary participation on budget usefulness).

Yet the participation of employees may go beyond the setting of budgetary targets per se, extending to processes for developing and implementing management accounting systems (De Haas and Kleingeld, 1999; Eldenburg et al., 2010). Considering performance measurement systems (PMS) specifically, employees may be involved in and have influence on a panoply of factors, including: the conceptualization of performance measures, defining the measures, identifying required data, adapting IT-systems, designing graphs and tables for the presentation of the measures, and even producing the periodic performance reports. There are only a few studies in management accounting that have investigated such a broader notion of participation in the development and implementation of performance measurement systems, and these generally found beneficial effects (i.e., Abernethy and Bouwens, 2005; De Haas and Algera, 2002; Hunton and Gibson, 1999; Kleingeld et al., 2004; Li and Tang, 2009; Wouters and Wilderom, 2008).

Investigating participation in the development and implementation of PMS is valuable because so little is known about why performance measurement affects performance. Many studies have investigated relationships between performance measurement and organizational performance (e.g., Chenhall, 2005; Davis and Albright, 2004; De Geuser et al., 2009; Farrell et al., 2008; Grafton et al., 2010; Ittner et al., 2003; Kelly, 2010; Lee and Yang,

* Corresponding author at: University of Twente, Institute for Innovation and Governance Studies, School of Management and Governance, Department of Business Administration, P.O. Box 217, 7500AE Enschede, Netherlands. Tel.: +31 53 489 5178; fax: +31 53 489 2159.

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

[1] Present address: Karlsruhe Institute of Technology, Germany.

1044-5005/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mar.2012.01.001

2010; Malina et al., 2007; Said et al., 2003; Widener, 2006). These studies assume performance measurement affects the behavior of individuals within the organization, which in turn facilitates the achievement of organizational goals (Burney and Widener, 2007; Burney et al., 2009; Covaleski et al., 2003; Hall, 2008). However, detailed empirical investigations into how employee behavior mediates the relationship between PMS and performance remain scarce (De Leeuw and Van den Berg, 2010; Hall, 2010; Luckett and Eggleton, 1991; Webb, 2004).

This study focuses on participatory development of performance measures and a particular type of behavior, namely employee initiative. Employee initiative is an increasingly important part of contemporary job performance (Campbell, 2000; Crant, 2000; Frese and Fay, 2001a) aimed at achieving continuous improvements in operational work processes. We define *PM participation* as the substantial impact of one or more employees on the content of the performance measures by means of which one (in this study: a department) is measured. We define *employee initiative* as self-starting, pro-active, persistent and pro-company behavior of individual employees (Frese and Fay, 2001b). The central question of our study is: why is *PM participation* related to *employee initiative*?

This study investigates performance measurement at the operational level in the organization, where performance measures are quite specific to the operational processes (Franco-Santos et al., 2007; McKinnon and Bruns, 1992; Melnyk et al., 2004). We focus on enabling performance measures that are intended to facilitate the responsibilities of employees, rather than primarily as control devices deployed by senior management (Adler and Borys, 1996; Ahrens and Chapman, 2004; Free, 2007; Wouters and Wilderom, 2008). Employees know a great deal about operational processes and the data that are generated, making it important to use their knowledge to develop and implement performance measures (Masquefa, 2008). We do not investigate the use of performance measures for formal evaluation and incentive purposes.

We intend to contribute to the management accounting literature on performance measurement systems by using a psychological theory to investigate our research question. This is important because psychological theories may give more complete and valid explanations of performance measurement effects (Covaleski et al., 2003; Kleingeld et al., 2004), thereby extending the existing management accounting body of knowledge on performance measurement. The theory we use in this study (the theory of planned behavior) has not yet been applied to employee initiative behavior, but it has been used to explore and stimulate various other kinds of behavior, such as quitting smoking, using condoms, and using public transportation (Fishbein and Ajzen, 2010). We show *employee initiative* behavior can also be studied through the same theoretical lens. Using this theory contributes to the management accounting literature because it investigates motivational, social and cognitive variables at the same time, which most likely are the major behavioral effects resulting from participation (Jeong, 2006). Earlier management accounting research has included motivation and/or capability variables, but social effects have been less investigated. In sum, the present study intends to provide an overall explanation for why *PM participation* is related to *employee initiative* by investigating all three of these important mediating variables simultaneously.

A secondary contribution of this study lies in the report in substantive detail precisely how *PM participation* actually came about and was shaped. This kind of process has received scant attention in the accounting literature heretofore (Otley, 1999; Abernethy and Bouwens, 2005). We report on a 1-year field study in a beverage manufacturing company where we jointly developed performance measures with their maintenance technicians. Using action research makes it possible to richly describe how employees reacted before, during and after they participated in developing their own performance measures.

This study was conducted in order to develop a theoretical explanation for why *PM participation* is related to *employee initiative*, and to provide initial empirical support for it. We did this by using systematic combining—continually going back and forth between theory and data (Dubois and Gadde, 2002). However, for the sake of clarity, from the outset we structure the paper around the developed model, which provides a structure that helps to convey the theoretical and empirical insights gained throughout this study about the effects of participative development of performance measures.

This paper is structured as follows. In Section 2 we articulate the theory that supports our model, and in Section 3 we lay out our methodology. Section 4 presents the empirical results with regard to qualifying and refining our basic model. Section 5 discusses a range of implications and limitations of our overall account.

2. Theory

We define *PM participation* as the substantial impact of one or more employees on the content of the performance measures by means of which one (in this study a department) is measured. This may include any aspect of the performance measures distinguished by Neely et al. (2002): the name; the purpose; the target; the formula; the frequency of measuring; the source of data; and the responsibility. By actually participating in the development of performance measures, employees' ideas about performance measures are taken seriously (Nørreklit, 2000). The goal is manifestly practical—to make performance measures useful for the involved employees in their everyday work. Of course, participation will not be a completely autonomous affair. For example, there may be guidance in the form of strategic priorities, constraints regarding the timely availability of resources for this developmental process, and project deadlines that the employees have to consider. *PM participation* may provide positive effects to the organization if it creates better quality performance measures (Abernethy and Bouwens, 2005). Good measurement properties of performance measures (such as sensitivity, precision, and verifiability) can reduce costly management control issues (Moers, 2006).

PM participation is not the same as the interactive use of performance measurement systems, which has also been investigated empirically (e.g., Bisbe and Otley, 2004; Henri, 2006; Widener, 2007). In terms of the framework developed by Ferreira and Otley (2009), the interactive use refers to how managers and employees use an *existing* PMS in their communication, whereas *PM participation* is about how managers and employees work together to design and implement a new or modified PMS.

Employee initiative is somewhat comparable to the term "work-related motivation" that is more common in management accounting.[2] However, work-related motivation is rarely measured directly and is often focused on a non-observable, internal state of mind (see Birnberg et al., 2007, for an overview). For example, Hunton and Gibson (1999) examined the link between a construct similar to *PM participation* and work-related motivation. They measured motivation indirectly through "self-efficacy" and perceived "participation congruence." We are interested not only in this internal state of mind, but also in employee *behavior*.

The basis of our model is the theory of planned behavior (TPB; Ajzen, 1991; Fishbein and Ajzen, 2010) that is widely used in psychological research to address how people can be motivated to behave in certain ways. It has to date not been used to explain or predict *employee initiative*, but we determined it would be fruitful given its effective use in a wide range of fields (Fishbein and Ajzen, 2010) including management accounting (e.g., Hill et al., 1996), organizational behavior (e.g., Dunn and Schweitzer, 2005), and change management (e.g., Jimmieson et al., 2008). The TPB differentiates between motivational, social and cognitive variables. This classic distinction is also used in, for instance, Birnberg et al.'s (2007) overview of psychology theory in management accounting research. Most research so far—both inside and outside of management accounting—has included only one or two of these types of variables at the same time.[3] The present research contributes to the literature by including all three mediating behavioral variables simultaneously and therefore giving a relatively complete explanation for the relation between *PM participation* and *employee initiative*.

The TPB distinguishes three antecedents of any particular kind of behavior: *attitude*—people's evaluation regarding the behavior, *norm*—the extent to which people think that most people who are important to them, want them to behave in a particular way, and *control*—the extent to which people feel capable of performing the behavior (see Ajzen, 1991, for the complete theory). Because the terms "Norm" and "Control" have a different connotation for management accounting scholars, we will below use different equivalent terms that are more intuitive: *social pressure* and *capability* to take initiative, respectively.

According to the TPB, it is possible to change people's behavior when an intervention is directed at one or more of its antecedents (Ajzen, 2006a). Therefore, we examine if *PM participation* influences *attitude*, *social pressure*, and *capability* to take initiative, and if all TPB relations hold with *employee initiative* as the dependent variable (see Fig. 1).

2.1. PM participation and attitude to take initiative

In Hackman and Oldham's (1976) "job characteristic model" the *attitude* to take initiative depends upon three *psychological* states: (1) experienced meaningfulness of the work, (2) experienced responsibility for the outcomes of the work, and (3) knowledge of the results of the work activities (Fried and Ferris, 1987; Hackman and Oldham, 1976; Johns et al., 1992). *PM participation* may invoke these psychological states and thus increase *attitude* to take initiative. The first state (experienced meaningfulness of the work) is invoked if *PM participation* gives rise to and reflects something employees believe in (Latham, 2003). In this case employees when trying to reach the goals do not have to sacrifice self-interest for the greater good (Bono and Judge, 2003). Hence they are likely to put more effort into reaching the goals (Sheldon and Elliot, 1998).

The second state (experienced responsibility for the outcomes of the work) is an inherent consequence of *PM participation* because it gives employees a certain amount of autonomy (Hackman and Oldham, 1976). When people have an influence on something, they often tend to become involved in making it work because they will perceive its success or failure as their own success or failure (Vroom, 1995, p. 267). In line with that kind of identification, *PM participation* makes them more positive about the developed performance measures (Abernethy and Bouwens, 2005; Wilderom et al., 2007). They will thus perceive the measures as a credible resource, which of course makes them more likely to accept their output (Ilgen et al., 1979) and use them to improve their work (Luckett and Eggleton, 1991).

The third state (knowledge of the results of work activities) is likely to be affected by *PM participation* as well. Performance measures provide feedback, increasing the knowledge of the employees necessary to make decisions (Demski and Feltham, 1976; Sprinkle, 2003; Van Veen-Dirks, 2009). Since participatorily developed performance measures have fewer measurement errors and better fit the needs of the employees (Abernethy and Bouwens, 2005; Cavalluzzo and Ittner, 2004), feedback is more likely to be accepted (Ilgen et al., 1979; Luckett and Eggleton, 1991) and the employees' knowledge of the results of their work improves. In summary:

Proposition 1. *If employees participate in developing their own performance measures, their attitude to take initiative becomes more positive.*

2.2. PM participation and social pressure to take initiative

In addition to attitudinal gains, participation in developing the measures also seems to give social benefits, especially when speaking of group participation, as we do

[2] Work-related motivation as used within management accounting is usually conceptualized as consisting of four processes: (1) "arousal"—the stimulation or initiation of energy to act; (2) "direction"—where energy or effort is directed; (3) "intensity"—the amount of effort expended per unit of time; and (4) "persistence"—the duration of time that effort is expended (Birnberg et al., 2007, p. 119).

[3] An exception is Erez and Arad (1986) who studied all three factors simultaneously. Their dependent variable was "performance," and they found that a combination of the three types of variables was indeed the best predictor.

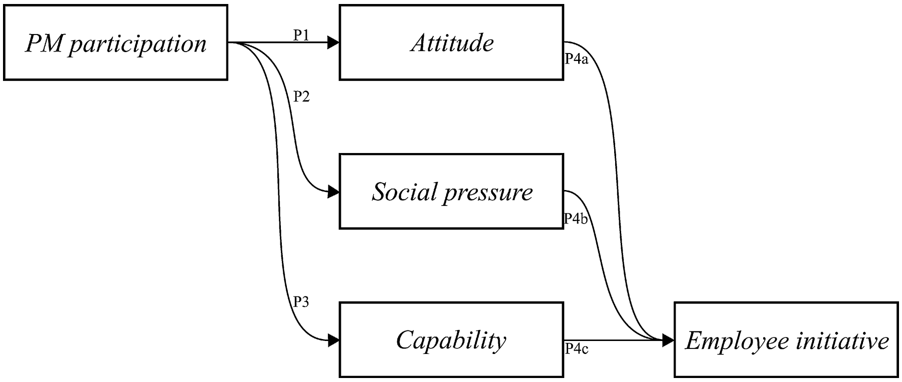


Fig. 1. Proposed model of the study.

in this study (Erez and Arad, 1986). We think that *PM participation* leads to more *social pressure* because performance measures can prioritize behavior (Collins, 1982; Sprinkle, 2003) and clarify the requirements of someone's work role (Hall, 2008). They indicate where employees should direct their effort, and the accompanying targets show how much effort they should put into it. After developing the performance measures together with their colleagues, employees are more likely to feel that they have to justify their performance, including the initiatives towards reaching the targets. Although these relations may also apply to non-participatory performance measures, it appears their influence is more prominent with self-developed performance measures. Acceptance of the measures is assumed to depend on the amount of influence someone has had on the selection and development of these measures (Luckett and Eggleton, 1991). A target should be accepted by the people concerned before it will have an effect on their behavior (Erez et al., 1985). We therefore propose:

Proposition 2. *If employees participate in developing their own performance measures, they feel more social pressure to take initiative.*

2.3. PM participation and capability to take initiative

Building on the ideas of enabling formalization (Adler and Borys, 1996; Ahrens and Chapman, 2004), *PM participation* is found to lead to performance measures that are perceived as enabling or empowering (Chiles and Zorn, 1995; Hall, 2008; Quinn and Spreitzer, 1997; Spreitzer, 1995, 1996). Enabling performance measures are perceived by employees as facilitative for their work, rather than as just a monitoring device for managers, as performance measures are often seen (Wouters and Wilderom, 2008). There are two mechanisms that may explain why employees feel more capable to take initiative if they have developed their own performance measures. The first derives from the literature on the cognitive mechanisms that explain the relation between participation and performance (e.g. Shields and Shields, 1998). It is argued that an important feature of the participatory process is the discussion that takes place between the employees and their leader. Due to these discussions people know better what to do and how to do it, making the performance measures more useful (Kleingeld et al., 2004) and giving the employees more actual and perceived capability.

PM participation may also affect *capability* via the decision-facilitating role of these developed performance measures. Individuals' knowledge and ability to make better decisions can be improved by providing feedback (Sprinkle, 2003), and accurate performance measures are providers of such feedback (Demski and Feltham, 1976; Sprinkle, 2003; Van Veen-Dirks, 2009). It is generally accepted that *PM participation* leads to performance measures of better quality (Abernethy and Bouwens, 2005; Cavalluzzo and Ittner, 2004), a key factor often leading to more self-efficacy with regard to reaching goals (Webb, 2004). Hence, we propose that *PM participation* makes employees more capable of taking initiative.

Proposition 3. *If employees participate in developing their own performance measures, their capability to take initiative increases.*

2.4. TPB antecedents and employee initiative

The theory of planned behavior advances the case that an individual's intention to perform a certain behavior depends on one's attitude, felt social pressure, and/or felt capability to perform the behavior; and that intentions are usually good predictors of behavior. Support for these relations is found in numerous studies and meta-analyses of diverse kinds of behavior (Fishbein and Ajzen, 2010). We foresee similar links with respect to *employee initiative* behavior and will below explain the rationale behind these propositions. We refer to Fishbein and Ajzen (2010) for the complete theory, and to the empirical papers that document relations that resemble those between *attitude* and *employee initiative* (Frese and Fay, 2001a,b; Fuller et al., 2006; Parker et al., 2006); *social pressure* and *employee initiative* (Crant, 2000; Frese and Fay, 2001a,b); and *capability* and *employee initiative* (Axtell and Parker, 2003; Morrison, 2006; Parker et al., 1997, 2006; Spreitzer, 1995; Thomas and Velthouse, 1990).

The relation between *attitude* to take initiative and actually taking initiative is intuitively reasonable if you consider the definition of *employee initiative*: it is practically impossible to be self-starting, pro-active and persistent if you do not feel positive about taking the initiative. The relation between *social pressure* and *employee initiative* exists because people generally fear the negative consequence of being different (Brehm et al., 2002). Finally, even if employees want to take initiative and feel the social pressure to do so, they may not actually take initiative if they do not feel capable of it (Fishbein and Ajzen, 2010). Taking initiative "requires the expectation of being in control of the situation and of one's actions" (Frese and Fay, 2001a, pp. 155).

Proposition 4a. *Employees' attitude to take initiative is positively related to employee initiative behavior.*

Proposition 4b. *Employees' felt social pressure towards taking initiative is positively related to employee initiative behavior.*

Proposition 4c. *Employees' capability to take initiative is positively related to employee initiative behavior.*

3. Method

3.1. Research design

This study is designed as action research, or more precisely as clinical field work (Baskerville and Wood-Harper, 1998), which means that the action researcher is involved with an organization in a helping role (Schein, 1987). The main action researcher worked 3 days a week on average at the site to do the clinical field work, and spent the other 2 weekdays at the university concentrating on the scientific part of the study. We chose action research because the research question concerns "understanding the process of change or improvement" (Coughlan and Coghlan, 2002, p. 227). Our research design was chosen in order to optimize the opportunity to gain valuable insight into how an organizational phenomenon as *PM participation* actually works in practice (Coughlan and Coghlan, 2002). Designing and conducting research in real-world settings improves the exchange of knowledge between researchers and practitioners (Anderson et al., 2001; Miller et al., 1997; Rynes et al., 2001; Van de Ven and Johnson, 2006), and if properly conducted can make accounting research more relevant in practice (Kasanen et al., 1993).

The intended contribution of the paper is to extend the current body of management accounting knowledge concerning the question of why *PM participation* is related to *employee initiative*. We did this by means of systematic combining: continually going back and forth between theory and data (Dubois and Gadde, 2002). From the beginning the research question was clear and we intended to answer it by using a psychological theory. We gradually focussed on the theory of planned behavior because it includes motivational, social *and* cognitive type variables, all relevant to adequately explaining the link between employee participation and performance (cf. Jeong, 2006). Meanwhile, working in concrete, everyday contexts gave us a better feeling about what actually goes on when performance measures are being developed together with employees. This experience helped us to gradually see more and more connections between these observations and existing literature, which enabled us to extensively embed our observations in theory. Although the study was undertaken for purpose of theory development, we used the opportunity to do some theory testing as well. Our qualitative study suggested that all three TPB variables seemed relevant to increasing *employee initiative*. Hence, at the tail end of the study we asked the employees to complete a questionnaire that would help us examine whether some of these relations were also statistically significant.

We designed this study in ways that adhered to Baskerville and Wood-Harper's (1998, pp. 103–104) seven validity criteria for action research: "(1) the research should be set in a multivariate social situation. (2) The observations are recorded and analyzed in an interpretive frame. (3) There was researcher action that intervened in the research setting. (4) The method of data collection included participatory observation. (5) Changes in the social setting were studied. [. . .] (6) The immediate problem in the social setting must have been resolved during the research. (7) The research should illuminate a theoretical framework that explains how the actions led to the favorable outcome."

The first five criteria are met through our choice of the research setting that we will describe in Section 3.2. Most interesting and relevant here are Criterions 6 and 7. To meet Criterion 6 the intervention should actually lead to more *employee initiative*. If it fails to lead to more *employee initiative* then it is impossible to examine how and why employees took more initiative after the intervention, so it would make the research invalid. In Section 4 we show that employees indeed eventually did take more initiative. Moreover, Criterion 7 can be read as suggesting this study illuminates a theoretical framework that explains why our intervention led to more *employee initiative*. This of course is our main research question and what our paper is all about. The developed theory is brought forward in Section 2, and in Section 4 we discuss how this model actually worked in the company in our case study.

In order to make our research replicable, we turn next to a very precise description of our methodology (see Checkland and Holwell, 2007). We start with a sketch of the research context that will help in the interpretation of the results. In Section 3.3 we describe each of the steps that we took to develop the performance measures together with the employees. In Section 3.4 we report how we captured the data and how we went about our analyses.

3.2. Research context

3.2.1. Organization

The organization under study is a medium-sized Dutch company in the beverage manufacturing industry. We focused on its maintenance department for the bottling lines. Fig. 2 shows the relevant part of the organizational chart. The director of the supply chain department was a member of the board of directors. The supply chain department consisted of five sub-departments, one of which was the bottling sub-department. The head of bottling was part of the "supply team" which met at least monthly to discuss the broader picture of the supply chain department. The supply team consisted of the supply chain director, the head of supply chain control, and the heads of the sub-departments of supply.

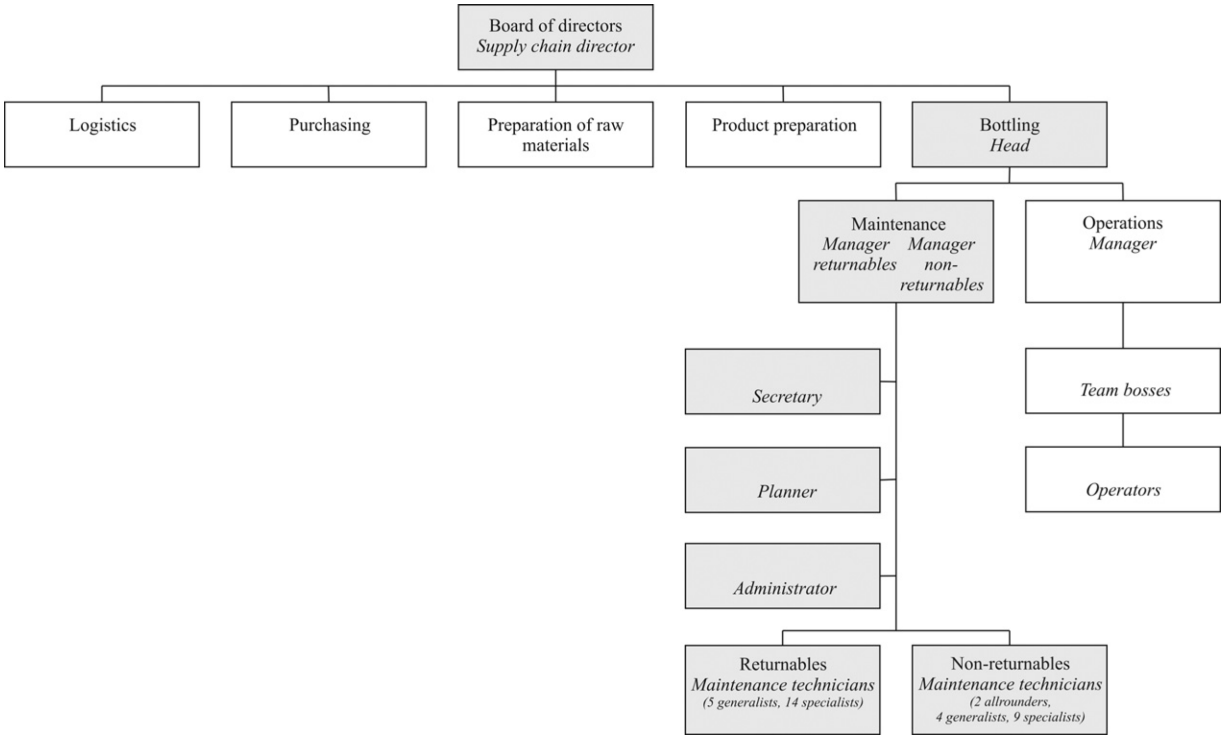


Fig. 2. Part of the company's organization chart including the stakeholders of the study.

The organizational chart changed slightly during our study, but the bottling sub-department was basically comprised of (a) the operators who were led by their own team bosses; and (b) the maintenance technicians who were led by two maintenance managers. Our study was situated among all the maintenance technicians and their managers. Of the 34 maintenance technicians, 16 were electro-technical and 18 were mechanical technicians. The remaining staff of the maintenance department included a planner, administrator, and secretary.

The bottling department has eight bottling lines. Each maintenance *manager* was responsible for four lines: one for lines that bottled using *returnable* materials, and the other for the lines using *non-returnable* materials. The processes of returnable and non-returnable materials differ because non-returnable materials are quality-checked before they enter the company, whereas returnable materials are not, which preempts directly comparing one-to-one the maintenance managers' performance. The maintenance *technicians* had an individual area of responsibility: 8 were responsible for one of the bottling lines, 24 for one kind of machine, and 2 were jack-of-all-trades and helped wherever and whenever they could.

Apart from the secretary all the employees of the maintenance department were male. The maintenance managers had both completed higher-level vocational education. One had been with the company for 28 years and had a departmental tenure of 20 years. The other, in contrast, had only recently joined the company at the beginning of our study. Four maintenance technicians had a lower-level and 30 had an intermediate vocational education background. The mean age of 33 of the 34 maintenance technicians was 45; their mean organizational tenure was 19 years. On average, they had spent 16 years working in this very same maintenance department.[4]

3.2.2. Changes over time

Besides our intervention, other relevant changes inside and outside the company were going on during our study. To put these changes into perspective, we refer to Fig. 3 that gives an overview of the study's timeline. We already mentioned that a new maintenance manager entered the company close to the beginning of the study. Moreover, in February 2008, the company was acquired by a larger, global, foreign based beverage manufacturing company. This new faraway owner had a decentralized structure in which the production locations work independently, and it seemed at the time the take-over would have no major consequences for the supply department. Nevertheless, in October 2008 a company-wide reorganization was announced and 10% of all the employees would lose their jobs. Within the maintenance department, about 11 of the 39 positions would disappear. By the end of our study, three technicians had taken early retirement and two technicians and the secretary had been transferred to other departments.

[4] One of the participating maintenance technicians did not provide information on age and tenure.

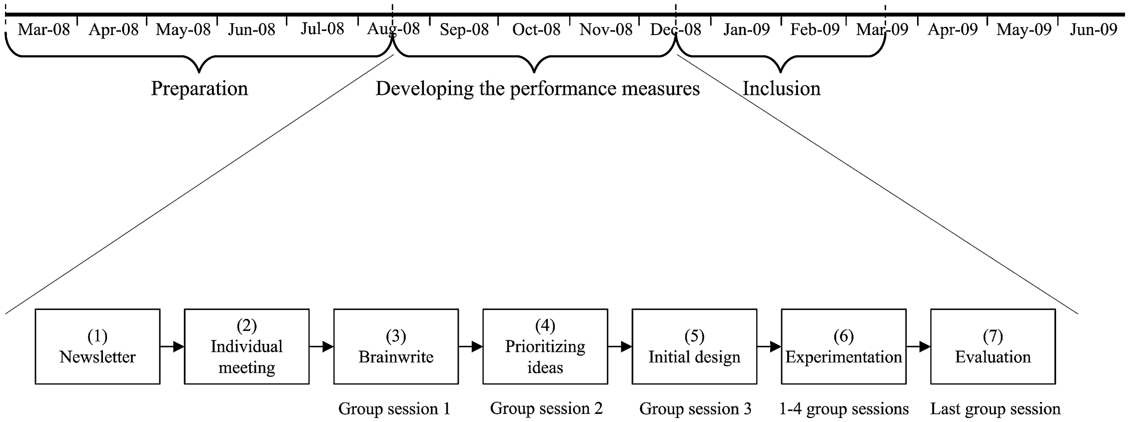


Fig. 3. Timeline of the study.

3.3. *Process*

The actual process of developing the performance measures—illustrated in Fig. 3—took 4 months. The rest of the 16 months of the study were used to prepare this process, to include the developed performance measures in the departmental routines and to collect data.

3.3.1. *Preparation*

The preparation consisted of several introductory meetings with several internal stakeholders. Moreover, four groups were formed. These groups were as diverse as possible, mixing the maintenance technicians from different lines and specializations. To make sure that the performance measures were explicitly in line with the goals of the organization, the head of bottling attached themes to the groups: (1) energy use, (2) material losses, (3) planned maintenance, and (4) machine failures. As part of the supply team, he had specific insights about the strategic priorities of the company and of the supply chain department. He wanted these four themes adopted because they were currently important for the bottling department in supporting the company's strategy. The rest of the supply team agreed with these themes.

3.3.2. *Developing the performance measures*

Fig. 3 summarizes the seven phases of the developmental process. It should be noted that in practice the transitions between the phases were more gradual than the schema suggests. Each phase can be briefly encapsulated as follows:

(1) Before the summer-break of 2008 a newsletter was e-mailed to all members of the maintenance department with information about the purpose of performance measures and the process that was going to be used to develop them. The technicians were asked to attend an individual meeting with the action researcher after the summer-break. We emphasized in the original newsletter and afterwards that the measures were supposed to assist *them* (the technicians) in improving their own work, rather than being used by management to evaluate their performance.

(2) During the individual meetings, the maintenance technicians could (a) explain the current ways of working in the maintenance department, (b) articulate their expectations about the project, and (c) ask questions about it. The meetings were also conducted to collect interview data.

(3) Each group created performance measures in five to eight group sessions led by the main action researcher. During each group's first session one of the two maintenance managers explained the importance of the project and the technicians participated in a so-called brain-write (e.g., Terhürne, 2008; Thompson, 2003). Somewhat analogous to brainstorming, they were asked to individually write down as many improvement ideas as they could think of for the theme of their group. After 10 min they handed their notes to their neighbors who used these to identify new or related ideas. This last step was repeated until everyone had received and elaborated upon the notes of everyone else. By beginning with improvement ideas rather than performance measures we had hoped to generate more efficient discussions and more commitment because: (a) it made the discussion immediately more concrete since improvement ideas are more tangible for the technicians than are performance measures, and (b) it showed the link between performance measures and taking initiative.

(4) The action researcher prior to each second group session categorized the improvement ideas and discussed them with the maintenance managers. During the second session the group prioritized and discussed them, selecting three areas within which they were going to develop performance measures.

(5) At the next session the action researcher helped the maintenance technicians to decide on the contents of the performance measures. She explained established criteria for making useable performance measures based on the Neely et al.'s (2002) performance measurement record sheet. This helped the maintenance technicians to specify the performance measures' purpose; relation to company goals; target; formula; data source; frequency of updating and discussing; and responsibility for updating, etc.

(6) The action researcher created a prototype of each performance measure before the subsequent session, and updated it before every next group session. The rationale for using prototypes was to have a more concrete discussion and make the measures as valid, reliable and understandable as possible (Wouters and Roijmans, 2011). The prototypes were based on information received during the sessions with the maintenance technicians and from others in the company, primarily those responsible for various information systems. The prototypes contained real data that were already being measured by the company's information systems.
(7) During the last group sessions, each group evaluated the developmental process and the results.

During the developmental process, the action researcher had regular meetings with the two maintenance managers where process and content issues were raised and addressed. Furthermore, with the same aim formal evaluation sessions took place before, during, and after the intervention with the maintenance managers, the head of bottling, and the head of supply chain control. The action researcher also kept the director of the supply chain department informed about the progress and results. These meetings helped the researchers to find solutions for context-specific problems during the process. Moreover, they enabled the managers to be alert about the progress of the process and be sure the technicians would work on strategically relevant performance measures. As it turned out none of these managers felt it was necessary to change the intervention process at any point in time.

3.3.3. Inclusion

All maintenance managers and technicians agreed to discuss the newly designed performance measures at least monthly during one of their daily line meetings. A daily line meeting is a half-hour morning meeting of the maintenance technicians that are present at the time, their manager, and the team boss of their bottling lines. At these meetings they discuss events of the past 24 h, as well as other issues related to the work of the maintenance technicians. The researcher joined some of the daily line meetings in which the performance measures were discussed. During these meetings, she helped the maintenance technicians explain the measures to others who had not participated in the making of a specific measure. These early morning meetings afforded the researcher with an excellent opportunity to see how the measures were being used, and what initial effects they seemed to be having.

3.4. Data collection and analysis

We used multiple data sources for our analyses. We collected qualitative data from all the meetings, observations and semi-structured interviews and relevant quantitative performance data from the company's information systems. Moreover, the maintenance technicians completed a questionnaire after the performance measures were in use.

The level of analysis in this study was the individual. We were interested in the participatory development process that individual employees experienced, and the effect this had on the employee initiative behavior of individuals through *attitude*, *social pressure* and *capability*. These variables were all at the individual level (see our model in Fig. 1). The process led to the development of aggregated departmental performance measures as well, but this is not part of our model.

3.4.1. Meetings and observations

Most of our qualitative data was gathered at 190 meetings with 96 different company employees. These sessions lasted approximately 200 h in total (see Table 1). The action researcher routinely took notes and made a report of each meeting, objectively noting date, starting time, duration, attending employees, attending researchers, the involved department, subject, reference to input for the meeting, reference to meeting notes, reference to company documents received, and type of contact (e.g., scheduled or ad hoc).

The notes were systematically coded in terms of "performance measurement," "attitude," "social pressure," "capability," "employee initiative," and "performance." In other words, all text relating to one or more of these constructs was highlighted and tagged with the name of the associated construct. Moreover, for each variable of interest the corresponding pieces of coded text were assembled in a separate listing.

3.4.2. Interviews

34 of the 190 meetings were semi-structured individual interviews with the maintenance technicians about *attitude*, *social pressure*, and *capability* to take initiative. Each interview began with an introduction aimed at putting the respondents at ease, explaining the aim, content and estimated duration of the interview. The scientific goal of the data collection was stressed. The technicians were told a project would start later that month in which the action researcher would help them develop their own performance measures. They were told that the final purpose of the project was helping them take more initiative in improving the performance of their department. The working definition of "initiative" was explained, and reminders of this definition were also given later in the interview.

Based on Ajzen's (2006b) and Francis et al.'s (2004) manuals for constructing TPB questionnaires, *attitude*, *social pressure* and *capability* to take initiative were measured directly with these questions: "What is your opinion about taking more initiative?" "What would colleagues think of you if you were always the one that came up with improvement ideas?" and "Do you think you are able to take initiative?"

Furthermore, questions were asked about the behavioral, normative, and control beliefs of the maintenance technicians. The answers gave us more and richer information about the contextualized meaning and examples of *attitude*, *social pressure* and *capability* and gave us a qualitative basis for assessing whether *PM participation* had influence on *attitude*, *social pressure* and *capability*. We asked the technicians about (1) their views on the advantages and disadvantages of taking initiatives; (2) the groups or persons that are explicitly positive or negative when coming up with and implementing improvement ideas;

Table 1
Specification of the meetings that the action researcher arranged and/or attended.

Phase	Activity	Number of meetings	Total time (in hours)	Number of different employees involved	Average number of employees per meeting
Preparation	Introduction	35	31	39	1.3
	Discussing project design and (preliminary) results	41	44	24	2.4
Intervention	Interviews with maintenance engineers	34	39	34	1.0
	Intervention sessions	27	33	32[a]	4.0
	Team: energy use	*6*	*5*	*8*	*4.4*
	Team: material losses	*8*	*9*	*8*	*4.3*
	Team: planned maintenance	*5*	*9*	*7*	*3.4*
	Team: machine failures	*8*	*10*	*9*	*3.9*
	Seeking information for specific PIs	28	18	22	1.3
Inclusion	Daily line meetings	7	4	26	5.7
Other	Other	18	29	All	Many
	Total	190	198	96	2.2

[a] The number of employees in the intervention sessions does not add up to 34, because one of the maintenance engineers was transferred to another department, and another one never showed up.

and (3) the factors or conditions that hinder or facilitate the spotting and implementing of improvement ideas (see Ajzen, 2006b; Francis et al., 2004). The responses to these questions indicated that, for example, *attitude* depends on whether taking initiative is perceived as a natural part of the job, the enjoyment or fun experienced, earlier experiences with improvement initiatives, and the appreciation received for taking the initiative.

As advised by Strauss and Corbin (1990) we began the analyses of the interviews with "open coding" giving every statement of the maintenance technicians a label. Then we classified the labels under "attitude," "social pressure" and "capability". Subsequently we selected and combined the labels into the aspects listed in Table 2. We recoded the interview texts using "attitude," "social pressure" and "capability" as codes so that we could assess if each respondent had given a response on each of those aspects, and if so whether it was positive, neutral or negative (see Table 2).

3.4.3. *Quantitative departmental performance data*

It is important to stress here that all the performance measures taken in this study refer to departmental performance rather than the performance of any of the individual maintenance technicians. The technicians developed and implemented five performance measures: (1) rejection due to under-filling, (2) rejection of empty bottles, (3) use of water, (4) use of electricity, and (5) use of compressed air. The first two measures were developed by the group "material losses", and the other three by the "energy use" group.[5]

These performance measures are directly related to the company goals for the bottling department: "cost reduction," "sustainability," and "efficiency improvement," as illustrated in Table 3. We use the results from the developed performance measures to assess the change in performance of the department. It was possible to reconstruct the measures for the period before the performance measures were developed (in the period June 2008–May 2009) because the measures are based on information already present within the IT-systems of the company.

3.4.4. *Questionnaire*

In June 2009, 25 maintenance technicians completed a questionnaire measuring *attitude*, *social pressure* and *capability* to take initiative, and employee initiative itself (see Appendix A). To measure *employee initiative*, we used Frese and Fay's (2003, p. 14) often used and thoroughly validated items. We used a 7-point Likert scale with anchors "totally disagree–totally agree." Earlier studies reported Cronbach's alphas of .80 (Frese et al., 1997) and .92 (Den Hartog and Belschak, 2007). In the present study Cronbach's alpha was .79.

Attitude, *social pressure* and *capability* to take initiative were each measured by four items (again using a 7-point Likert scale with anchors "totally disagree–totally agree") that were constructed following Francis et al. (2004). Cronbach's alphas were .91 for *attitude* and .66 for *social pressure*, but only .20 for *capability* to take initiative. In hindsight, we concluded two items that measured *capability* did not really measure what we had intended. Deleting them increased Cronbach's alpha to .36, which of course was still unacceptably low. Since there was no better alternative, we nevertheless used this scale. As a robustness check we also performed all analyses with the single item that best represents the *capability* construct (*I am confident that I could*

[5] It was not possible to develop performance measures with the other groups ("planned maintenance" and "machine failures") mainly because the IT-system was not capable of generating such measures, and higher management did not want to invest in adjusting the extant IT-system. This does not mean that these themes were irrelevant for top management. Managers repeatedly told us these themes were vital to the organization. They already had one employee working on defining the requirements of such an IT-system (for managerial purposes) before our project started. But that project was cancelled after the take-over when the company was not allowed to make such investment decisions in the remaining time of the study. We focused on the effects of those performance measures that were put into practice, rather than the ones that were not, to better understand why *PM participation* can lead to more *employee initiative*. We refer to Bourne and colleagues (Bourne, 2005; Bourne et al., 2002) for more information about why some performance measurement initiatives succeed and others do not.

Table 2
Results of the interviews.

Variable	Aspect	Response							
		Positive[a]		Neutral[b]		Negative[c]		No response[d]	
Attitude		29		4		1		–	
	Part of the job		20		–		2		12
	Fun		16		–		2		16
	Experience		12		1		3		18
	Appreciation		2		6		16		10
Social pressure		23		9		2		–	
	Maintenance technicians		9		5		3		17
	Operators		18		2		2		12
	Managers		14		–		–		20
	The company		17		–		–		17
Capability		30		2		2		–	
	Knowledge, skills, abilities		10		1		1		22
	Opportunity		11		–		10		13
	Facilitation by the manager		6		2		10		16
	Time		5		1		18		10
	Money		8		1		15		10
	Communication and cooperation		11		5		14		4

[a] The numbers in this column indicate the number of maintenance technicians that mentioned that they perceived the variable/aspect concerned as being present.
[b] The numbers in this column indicate the number of maintenance technicians that mentioned that they perceived the variable/aspect concerned as being not explicitly present or absent.
[c] The numbers in this column indicate the number of maintenance technicians that mentioned that they perceived the variable/aspect concerned as being absent.
[d] The numbers in this column indicate the number of maintenance technicians that did not mention the aspect.

Table 3
Contribution of the performance measures to the goals of the company.

Company goal	Performance measure	Why?
Cost reduction	Rejection due to under-filling	Less loss of product
	Rejection of empty bottles	Less loss of bottles
	Energy use	Less costs of energy
Sustainability	Rejection due to under-filling	Less waste
	Rejection of empty bottles	Less waste
	Energy use	Less use of energy
Efficiency improvement	Rejection due to under-filling	Less rejection of products leads to a higher efficiency
	Rejection of empty bottles	Less rejection of bottles leads to a higher efficiency
	Use of water	Less water is wasted when the lines are functioning better

think up and carry out improvement ideas by myself). In all other measures the scale scores were based on the average of the items.

4. Results

4.1. Results of the model

Propositions 1–3 are based on both qualitative and archival data. They state that *PM participation* affects the three TPB variables (*attitude*, *social pressure*, and *capability* to initiative). We investigated whether the maintenance technicians improved on these variables and on the developed performance measures. As a reference point, Table 4 gives index numbers of the production level of each line per month. Propositions 4a–c are examined based on the questionnaire data.

4.1.1. PM participation and attitude to take initiative

4.1.1.1. Attitude before. At the outset of the study, several people in the company felt the maintenance technicians' work attitudes would be quite negative, mainly because they had been subjected to several failed organizational changes in recent years. Amidst this skepticism, the manager of the bottling department was clearly perplexed in stating that "everyone has good intentions, but somehow improvement is not achieved." These good intentions were confirmed in nearly all the interviews held before the performance measures were developed: 29 out of 34 technicians said they felt positive about taking an initiative, 4 were neither positive nor negative, and only 1 was negative about it (see Table 2).

We divided the technicians' responses to the interview questions regarding *attitude* into four aspects: "part of the job," "fun," "experience" and "appreciation" (see Table 2). Many technicians during the interview noted they already considered improvement as part of their job, and some of them explicitly stated they liked it, or they had had earlier positive experiences with improvement efforts. Nevertheless, at the same time many complained about the lack of appreciation they received from management. "We only hear from them when we have done something wrong" was

Table 4
Production June 2008–May 2009 (index number)[a].

Responsible manager	Production line	Jun-08	Jul-08	Aug-08	Sep-08	Oct-08	Nov-08	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09
Returnable	1	100	56	73	64	58	52	67	28	69	60	76	60
	2	100	120	124	115	89	111	139	52	37	24	86	101
	3	100	75	76	104	58	50	78	95	81	96	105	108
	4	100	139	124	137	79	52	121	74	74	53	75	82
Non-returnable	5	100	100	25	33	30	126	37	20	34	15	26	111
	6	100	7	2	20	19	23	5	13	11	10	23	11
	7	100	136	53	70	65	47	102	43	59	75	61	101
	8	100	99	51	66	55	62	44	59	51	73	76	86

[a] This table makes it possible to compare the performance of the performance measures with the production level. For confidentiality reasons, we use index numbers instead of real production numbers. The production of June 2008 forms the base value. The numbers express the ratio of that month's production level to the base value.

Table 5
Results of "rejection of under-filled bottles" (% in relation to production).

Responsible manager	Production line	Months before measure was finished					Months after measure was finished						
		Jun-08	Jul-08	Aug-08	Sep-08	Oct-08	Nov-08	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09
Returnable	2[a]	0.2%	0.3%	0.2%	0.3%	0.2%	0.3%	0.4%	0.6%	0.6%	0.6%	0.5%	0.5%
	3	0.2%	0.2%	0.4%	0.2%	0.1%	0.2%	1.0%	0.3%	0.4%	0.3%	0.2%	0.2%
	4[a]	0.7%	0.4%	0.5%	0.4%	0.6%	0.2%	0.2%	0.2%	0.2%	0.4%	0.2%	0.2%
Non-returnable	8	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%

[a] The T-test shows a significant change since implementation ($p < .01$): line 2: $T(7) = -4.89$; line 4: $T(7) = 4.94$.

a common sentiment. This was corroborated by the interim manager of the maintenance department who repeatedly said: "The motivation of the maintenance technicians to come forward with improvement ideas is decreasing more and more, because they never get feedback on the results of their ideas." Thus, any improvement in employees' *attitude* should be visible in the "appreciation" aspect.

4.1.1.2. Attitude after. In November 2008, when most of the performance measures had already been implemented, the maintenance managers mentioned: "The maintenance technicians now talk to each other about the performance measures and about what could be improved." A month later one of the maintenance managers reported that the technicians were actually checking the results of each performance measure update. Moreover, during the daily line meetings, both the action researcher and the maintenance managers noted that the maintenance technicians now seemed to be focused more on improving than before. Example 1 shows the most prominent case of improved *attitude* during the development of the performance measures.

Example 1—rejection of under-filled bottles on bottling line 4. In October 2008 the maintenance technicians reviewed the output from the first version of the performance measure "rejection of under-filled bottles" (see Table 5 for this performance measure's data) in which all but one bottling line had a rejection percentage of about 0.2% or lower. Line 4 was the exception—it had a mean rejection percentage of about 0.5% ($SD = 1.1 \times 10^{-3}$) from June through October. The technicians of that line were shocked and aimed to lower that percentage to 0.2%. They became eager to improve this percentage after seeing the current performance of the other lines, so the next month they revised their line. The mean rejection percentage was indeed on average 0.2% ($SD = 0.9 \times 10^{-3}$) for the next 7 months (November through to May), a statistically significant improvement ($T(7) = 4.94$; $p < .001$). In March 2009 the percentage rose, but this problem was quickly resolved without any interference by the maintenance managers.

In this example, the maintenance technicians found it obvious that putting effort into improving the percentage was worthwhile. In contrast, in the following example the technicians did not see the benefit—at least not initially. They needed additional information about the costs before they were willing to make improvement efforts.

Example 2—use of water and compressed air on all bottling lines. The early versions of the performance measures regarding energy use did not immediately lead to better results. When in December 2008 the energy costs in the bottling department were made explicit, and known to everyone, the maintenance technicians were very surprised to learn that the total energy costs of the department were equal to that of at least 10 full time employees. The technicians thereafter commented that this financial aspect of the performance measure motivated them to improve. They stated they had simply not realized the scale of the benefit to be gained from improving that particular aspect of their work.

By February 2009 the use of water (see Table 6) improved. In the first 8 months the realized performance was .5% better than the target (SD = 6%). For the months February–May 2009 it was on average 18% (SD = 6%)

Table 6
Results of "use of water" (m^3).

Responsible manager	Production line[a]	Months before measure was finished							Months after measure was finished				
		Jun-08	Jul-08	Aug-08	Sep-08	Oct-08	Nov-08	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09
Returnable	1	820	585	692	629	1260	1621	1240	468	1956	1649	1024	796
	2	8804	9086	11,629	11,588	7460	7597	8064	5475	2628	3472	6727	4685
	3	7069	5431	5904	7965	5332	3940	8917	7377	5088	5734	5919	6082
	4	4746	6034	4935	4967	3176	2824	4116	3299	2668	1877	2455	2650
Non-returnable	5	256	199	145	197	165	244	155	144	136	140	157	232
	6	1131	314	384	331	235	159	142	86	84	91	134	88
	7	3291	4126	1535	2452	2238	1734	2348	1661	1957	2729	2873	3511
	8	5004	4500	2953	3049	2804	2606	1984	2197	2034	3468	3702	3856
Total		31,121	30,275	28,177	31,178	22,670	20,725	26,966	20,707	16,551	19,160	22,991	21,900
Target[b]		31,645	31,290	26,212	28,360	23,259	22,402	27,865	21,471	21,346	23,154	25,448	28,405
(Target-Total)/Target[c]		2%	3%	−7%	−10%	3%	7%	3%	4%	22%	17%	10%	23%

[a] To give more information about where the water was used for, besides totals, the performance measures also showed the use of water per production line.

[b] Use of water is only partly dependent upon production. To make the performance measure more precise and informative, the target consisted of a fixed part and of a variable part that was based on the production level. The fixed and variable parts were determined with linear regression analyses on the use of water over all months with data available until Nov-08 (=Apr–Nov).

[c] Before the performance measure was finished, the total use of water was on average 0.5% (SD = 6%) better than the target and after it was on average 18% (SD = 6%) better than the target. A T-test shows this is a significant improvement ($T(6) = -4.69$; $p < .01$).

better than the target: a statistically significant improvement ($T(6) = -4.69$; $p < .01$). This statically significant improvement did not extend to compressed air (see Table 7), primarily due to a defect in the bottling line machinery that resulted in a major negative result in April 2009 (−27% compared to the target). However, when we remove this outlier, there is a statistically significant improvement: in the first 8 months they were on average .1% (SD = 5%) better than the target, and in February, March and May 2009 they were 14% (SD = 5%) better than the target ($T(4) = -4.24$; $p < .01$).

In the evaluation sessions the maintenance technicians praised the fact that the newly developed performance measures allowed them to see how well they were doing their job. This gave them a feeling of appreciation, which was further reinforced when their managers also used the information from the performance measures to compliment them for their work. Before the performance measures were put into play such positive feedback had hardly ever been received. This indicates that the "appreciation" area of *attitude* had improved. In the section "attitude before," we claimed that this area of *attitude* needed the most improvement. These changes in patterns of behavior support Proposition 1.

4.1.2. PM participation and social pressure to take initiative

PM participation also increased *social pressure* to take initiative (Proposition 2). We will again examine this relation through the use of qualitative and archival data.

Table 7
Results of "use of compressed air" (nm^3).

Responsible manager	Production line[a]	Months before measure was finished							Months after measure was finished				
		Jun-08	Jul-08	Aug-08	Sep-08	Oct-08	Nov-08	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09
Returnable	1	214,981	203,748	200,485	189,455	192,755	180,442	183,726	176,161	163,406	174,011	425,468	212,364
	2	54,080	63,756	88,267	90,289	75,222	77,718	84,329	46,184	22,789	27,060	52,131	44,772
	3	52,215	36,983	36,183	55,203	43,663	35,691	61,060	57,422	38,953	42,519	47,449	47,927
	4	196,335	206,435	181,359	186,971	139,066	118,490	214,376	145,273	135,567	118,769	131,172	148,103
Non-returnable	5	10,630	11,324	9456	8969	10,948	11,399	11,163	3187	4621	5906	5265	8609
	6	n.a.	n.a.	12,670	12,269	7346	1626	912	816	772	742	1508	5930
	7	45,438	63,186	25,120	39,689	38,586	33,353	55,096	24,572	31,078	39,770	34,894	49,144
	8	1282	1313	645	923	768	869	609	820	580	906	1111	1264
Total		574,961	586,745	554,185	583,768	508,354	459,588	611,271	454,435	397,766	409,683	698,998	518,113
Target[b]		611,462	606,612	537,258	566,594	496,939	485,237	559,837	472,513	470,810	495,495	548,716	567,210
(Target-Total)/Target[c]		6%	3%	−3%	−3%	−2%	5%	−9%	4%	16%	17%	−27%	9%

[a] To give more information about where the compressed air was used for, besides totals, the performance measures also showed the use of compressed air per production line.

[b] Use of compressed air is only partly dependent upon production. To make the performance measure more precise and informative, the target consisted of a fixed part and of a variable part that was based on the production level. The fixed and variable parts were determined with linear regression analyses on the use of compressed air over all months with data available until Nov-08 (=Jul–Nov).

[c] Before the performance measure was finished, the total use of compressed air was on average 0.1% (SD = 5%) better than the target and after (leaving out the outlier of Apr-09 caused by a defect in line 1's meter) it was on average 14% (SD = 5%) better than the target. A T-test shows this is a significant improvement ($T(4) = -4.24$; $p < .01$).

4.1.2.1. Social pressure before. In the initial interviews we asked the maintenance technicians what they thought colleagues would think of them were they themselves to come up with improvement ideas. Out of the 34, 23 of them thought their colleagues would react positively (see Table 2), and the others said that should some colleagues react negatively it would not stop them from consulting with colleagues. We asked the maintenance technicians which groups or persons they thought would be *explicitly* positive or negative to the creation and the implementation of improvement ideas (as mentioned in Section 3.4.2). They mentioned other "maintenance engineers," "line operators," their "managers," and "the company," and they expected mostly positive responses (see Table 2). On the other hand, some could also think of negative responses from their fellow maintenance technicians and line operators: if the performance of the machines improves "too much" both line operators and maintenance technicians would have to fear for their jobs. Yet at the time of the interviews they had not thought this fear was realistic. In summary, most maintenance technicians felt that the social pressure was directed towards taking more initiative, some felt the social pressure was against taking more initiatives and some did not feel it at all. In other words, there was a broad mix of interpretations of colleagues' opinions regarding taking more initiatives.

4.1.2.2. Social pressure after. Our qualitative data suggest that the performance measures made it explicit that improvement was expected. The performance measures provided the maintenance technicians with a target that was developed together with people who are important to them. Consequently, it was a manifestation of *social pressure*. This target was an explicit goal in Example 2 above. However, even when no explicit goal was set, we did find instances where the performance improved after the performance measures were discussed during the daily-line meeting. Example 3 illustrates this and together with Example 2 supports Proposition 2.

Example 3—use of compressed air on bottling line 5. The performance measure "use of compressed air" (see Table 7) showed that bottling line 5 had used on average 10,556 Nm^3 (SD = 962) compressed air per month over the previous 7 months, despite rarely being in operation. When in December 2008 the maintenance manager and technicians discussed this at a daily line meeting they quickly concluded the strong discrepancy implied there were leakages. They all agreed they would try to find and repair them soon. Afterwards the amount of compressed air used by that line dropped significantly to an average of 5518 Nm^3 (SD = 2000) over the following 5 months ($T(5) = 5.22$; $p < .01$).

4.1.3. PM participation and capability to take initiative

Finally, we will discuss how *PM participation* helped to increase *capability* to take initiative (Proposition 3).

4.1.3.1. Capability before. Most maintenance technicians said in the interviews that they felt capable of showing initiative in their work (30 of the 34, see Table 2). Triggered by the question "are there any factors or conditions that hinder or facilitate you in finding and implementing improvement ideas?" they discussed several aspects of their work regarding their *capability* to take initiative. We summarized them as: "knowledge, skills and ability," "opportunity," "facilitation by the manager," "time," "money," and "communication and cooperation." The performance measures were expected to influence all of these aspects.

Initially, the maintenance technicians' "knowledge, skills and abilities" seemed to be operating satisfactorily (see Table 2). Many technicians said they usually had answers to the problems that arose in the bottling department, and if not they were generally confident *someone* would know a solution. According to the previous interim manager of the maintenance department the education and knowledge level of the maintenance technicians was good; and current maintenance managers said the technicians knew the bottlenecks in the lines better than anyone. Accordingly, many indicated that there of course was ample "opportunity" to improve (see next line in Table 2), also because they were of the opinion that a lot went wrong in the bottling department.

With regard to "facilitation by the manager," the maintenance technicians noted that their managers did not take enough time to assess and approve their suggestions. They could thus not carry out all the possible improvements they had in mind, because they needed permission before trying to implement an improvement idea. In a similar vein some maintenance technicians found it difficult to convince the management to invest "time" and "money" (see Table 2) in projects resulting from their improvement ideas. The frustrated technicians coped with this inattention in different ways—some went to the head of the bottling line, others to the maintenance managers, and others just ordered the materials they needed directly from the planner. This may explain why some technicians say there is enough time and money to implement their own improvement ideas, while others do not.

Maintenance technicians reported high levels of bureaucracy within the company, which made implementing improvement ideas difficult and time-consuming. Some technicians reported that they were often sent "from pillar to post," and eventually stopped trying. Other technicians stated that they did not always tell their managers about the improvement ideas they are implementing. This is a typical problem with regard to "communication and cooperation." In May 2008 the daily line meetings were introduced (see Section 3.3) which positively influenced the information transfer between the technicians and their managers, and vice versa.

4.1.3.2. Capability after. One of the maintenance technicians of bottling line 4 stated that the performance measures' most important contribution was that the technicians could finally demonstrate to the management the importance of improving the filler station of the bottling line. Consequently their manager was more supportive, allowing them to spend more "time" and "money" which helped them to decrease the rejection percentage due to under-filling (see Example 1 above). Thus, the aspects (Table 2) "support of manager," "time" and "money"

Table 8
Results of "use of electricity" (MWh).

Responsible manager	Production line[a]	Months before measure was finished							Months after measure was finished				
		Jun-08	Jul-08	Aug-08	Sep-08	Oct-08	Nov-08	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09
Returnable	1	26.5	17.8	18.3	16.2	16.2	15.8	15.7	9.2	14.0	18.5	23.5	12.0
	2	72.2	79.4	81.2	86.1	62.7	71.6	81.5	42.9	29.3	32.6	62.4	55.6
	3	82.6	66.2	69.8	87.8	59.4	45.5	81.4	74.8	61.1	72.0	74.3	79.9
	4	76.7	98.5	90.8	92.4	60.8	46.9	89.6	59.7	56.5	47.4	64.5	62.3
Non-returnable	5	2.3	1.9	0.6	2.0	1.4	2.3	0.9	0.2	0.7	0.5	0.8	2.3
	6	14.2	6.5	5.5	6.2	4.5	3.2	2.0	2.4	3.0	2.0	3.0	2.5
	7	93.2	119.9	56.0	75.4	73.9	55.2	87.9	50.3	62.1	82.9	73.8	84.8
	8	84.9	58.6	50.2	50.2	42.0	51.9	19.4	57.8	34.2	62.4	72.5	73.8
Total		452.5	449.0	372.4	416.3	320.9	292.3	378.3	297.4	261.0	318.3	374.7	373.3
Target[b]		459.8	453.8	367.4	403.9	317.2	302.6	395.5	286.8	284.7	315.4	381.7	404.7
(Target-Total)/Target		2%	1%	−1%	−3%	−1%	3%	4%	−4%	8%	−1%	2%	8%

[a] To give more information about where the electricity was used for, besides totals, the performance measures also showed the use of electricity per production line.
[b] Use of electricity is only partly dependent upon production. To make the performance measure more precise and informative, the target consisted of a fixed part and of a variable part that was based on the production level. The fixed and variable parts were determined with linear regression analyses on the use of electricity over all months with data available until Nov-08 (=May–Nov).

improved with the introduction of the performance measures.

"Communication and cooperation" improved somewhat with the introduction of the daily line meetings where both the maintenance manager and the maintenance technicians raised improvement ideas. Once the implementation of the performance measures began they started discussing improvement opportunities more routinely and in a structured manner, which further improved communication and cooperation in the maintenance department. Moreover, the development process itself led to more knowledge transfer between maintenance technicians. In the evaluation sessions, many technicians pointed with approval to the "discussions" during the sessions that "allowed them to learn from each other."

In general the process of developing performance measures gave the maintenance technicians more insight into their own improvement opportunities. Before they became involved in the development of their own performance measures, they were unaware so many improvements were possible. Although they knew a lot was going wrong in the maintenance department, they failed to accurately grasp what the problems were or how to solve them. The development process and the performance measures made them more competent to upgrade their overall performance. We see this change as supporting Proposition 3.

The next example, one in which the performance measures did not improve *capability*, may show that the *capability* to take initiative is a necessary condition for actually taking initiative.

Example 4—use of electricity on all bottling lines. In Example 2 we saw that the maintenance technicians managed to increase the performance with regard to the use of water and compressed air. The same group of technicians developed the measure for the use of electricity (Table 8). However, during one of the first meetings, the maintenance technicians mentioned that they had no influence over the use of electricity. They said that it was not up to them to implement all the electricity-use improvement ideas they had written down at the brain-write session. The intended performance measure had nevertheless been developed, but at the time the action researcher left the company the technicians were still unable to improve the situation.

4.1.4. TPB antecedents and employee initiative

The questionnaire data provide the basis for examining Propositions 4a–c. Table 9 shows the correlations between all variables, including many demographic variables. The significant correlations found between all TPB variables and *employee initiative* seem to support P4a–c (P4a: $r = .58, p < .01$; P4b: $r = .43, p < .05$; P4c: $r = .38, p < .05$[6]). Moreover, we find a significant correlation between *attitude* and *social pressure* to take initiative ($r = .68$, $p < .01$). Table 10 shows the results of the regression analysis used to determine which variables contribute most to the variance in *employee initiative*. Since we neither found any correlations between any of the demographic variables and any of the variables of the model, nor had a theoretical reason to expect such a relation, demographic variables should not be included in the regression specifications (Becker, 2005). The link between *capability* and *employee initiative* is the only factor that remains significant when all the variables are analyzed at the same time.

4.2. Influence of PM participation

In Sections 4.1.1–4.1.3 we have shown that the *attitude*, *social pressure* and *capability* to take initiative all increased after the departmental performance measures were implemented. A key question is: was the participatory nature of the intervention process important for this result, or would top–down development of the performance measures have

[6] When *capability* is only measured with the item that best represents the construct (*I am confident that I could think up and carry out improvement ideas by myself*), the significance levels are the same in both the correlation and regression analyses.

Table 9
Scale characteristics and correlations.

		α	Mean	SD	N	1	2	3	4	5	6	7	8	9	10
1	Responsibility area[a]		1.76	0.44	25										
2	Discipline[b]		1.60	0.50	25	.11									
3	Manager[c]		1.40	0.50	25	.08	.00								
4	Organizational tenure		16.4	11.3	25	−.16	.41*	−.48**							
5	Departmental tenure		13.7	10.4	25	−.11	.20	−.47**	.79**						
6	Age		43.2	9.21	25	−.10	.36*	−.28	.73**	.70**					
7	Education[d]		1.88	0.33	25	.08	−.30	.30	−.49**	−.21	−.25				
8	Attitude	.91	6.07	0.81	25	−.21	.04	−.04	.00	−.02	.04	.16			
9	Social pressure	.66	4.89	0.95	24	−.25	−.07	.16	.07	.00	.06	−.25	.68**		
10	Capability[e]	.36	5.24	0.89	25	.32	.27	.01	−.11	.04	.03	.10	.13	−.18	
11	Employee initiative	.79	5.64	0.62	24	−.11	.07	.09	−.07	.03	−.12	−.13	.58**	.43*	.38*

[a] 1 = responsible for one line; 2 = responsible for one kind of machine.
[b] 1 = electro-technical; 2 = mechanical.
[c] 1 = responsible for returnable; 2 = responsible for non-returnable.
[d] 1 = lower-level; 2 = intermediate level.
[e] Significance levels are the same if only one capability item is used.
* $p < .05$ (one-tailed).
** $p < .01$ (one-tailed).

generated the same desirable effects? The following example indicates that indeed participation did matter. It shows that the maintenance technicians—who were involved in the development process—took action when the performance in the measures decreased; whereas the responsible maintenance manager—who was not directly involved in the development process—did not take any action because he did not believe the numbers.

Example 5—rejection of under-filled bottles on bottling lines 2 and 3. When the performance measure "rejection of under-filled bottles" (Table 5) was made, the maintenance technicians of bottling lines 2 and 3 were convinced that their rejection percentage due to under-filling was already satisfactory. Yet about 1 month later, following changes made to bottling lines 2 and 3, the rejection percentages of these two lines began rising. Bottling line two's percentage rose because the line began to be used for small batches only, and batch changes are always followed by under-filling. The maintenance technicians were familiar with this and believed they were thus unable to reach the target again. Regarding bottling line 3, the maintenance technicians took action after recognizing the decreased performance on the measure was stable, leading them to believe the target could only be reached again if they themselves improved the bottling line. Just before the performance was satisfactory again, in March 2009, the responsible maintenance manager—who had not attended the sessions—saw the decreased performance on the measures. He stated he did not believe those statistics because he was (falsely) convinced that it was impossible to perform badly on under-filling and be satisfactory in terms of line efficiency at the same time. Participation in the development of the performance measure on under-filling seems to explain why the technicians felt they should improve, while the manager did not.

We have another indication that *PM participation* worked well in this setting. Initially, when we told some maintenance technicians that we were going to develop performance indicators together with them, they reacted negatively. Examples of their reactions are: "That is impossible for such a complicated process" and "I don't think we should be evaluated." The action researcher said that she would actively help them and that the resulting performance measures would only be used to facilitate them in their jobs. The maintenance manager who was present endorsed this process. Contrary to their earlier negative reactions, in the evaluation sessions after the performance measures were developed, these same technicians were now convinced of the value of using performance measures. They had come around to the idea the measures really showed how they performed and these positive results were a consequence of the specific process that was used. They especially liked the fact that the process was begun with them thinking of improvement ideas, because that made the performance indicators more prospectively relevant to them. Table 11 shows these and the other reactions during the evaluation sessions.

Although they were disappointed about not being able to realize their ideas, the maintenance technicians that were not allowed to implement their performance indicators were positive about the process. They said the process had helped them to understand what performance measures are and how to use them. Moreover, they valued the fact they were finally able to speak constructively to their colleagues in other parts of the department. Moreover, they were excited about the large number of improvement ideas that came up during the brain-write sessions.

Table 10
Results of the regression analyses predicting employee initiative.

Variable	B	SE B	β
(Constant)	11.86		
Attitude	0.49	0.32	.36
Social pressure	0.29	0.27	.26
Capability[a]	0.91	0.42	.38*

$R^2 = .46$; $N = 24$.
[a] Significance levels are the same if only one capability item is used.
* $p < .05$.

Table 11
Comments given during the evaluation sessions.

	Process	Result
Good as it was	Starting with improvement ideas The diversity of the group Stimulating discussion The structure of the sessions "Prototyping" Enthusiasm of the action researcher	Many improvement ideas Insight in advantages of performance measures We formulated goals More insight into costs More insight into effects of our work We are more critical of our work
Could be improved	Attendance percentage It is very time-consuming The time between sessions was too long Our managers should motivate us more Many ideas are outside of our influence	We are afraid the positive results will fade away

4.3. Quality of the measures

We think that the positive influence of *PM participation* on the behavior of employees partly occurred because involving employees leads to better quality performance measures. In terms of Moers (2006), quality consists of precision, sensitivity and verifiability of performance measures, which were all positively influenced by the participatory development process. Verifiability increased because the performance measures were based on sources that were identified by the maintenance technicians, so they knew exactly where the numbers originated. Moreover, discussions of prototypes sometimes led to better precision and sensitivity in the performance measures (see Example 6).

Example 6—use of electricity prototypes. The first version of the performance measure "use of electricity" was developed by the action researcher. It was based on the maintenance technicians' initial answers to the performance measurement record sheet, and conversations with a staff employee of the bottling department well versed in the information system that stores information about the use of electricity in the bottling department. The first prototype included every kind of electricity use the information system contained pertaining to the bottling department. When the prototype was discussed with the maintenance technicians at the next session they indicated that many of these identified electricity usage points were actually not part of the bottling department. These usage points were thus eliminated from the next prototype in order to make the measures more precise. Moreover, the maintenance technicians wanted to exclude the battery charging station of the forklift trucks, because this used a constant amount of electricity throughout all of the previous months. This narrowing of the energy use performance measure also increased the sensitivity of this measure.

Another way in which the quality of the performance measures increased is detailed in Example 7.

Example 7—use of water on bottling line 1. In the first week of December 2008, the maintenance technicians discussed the performance measures at a daily line meeting. They noticed the measures showed that the use of water on bottling line 1 had recently increased a lot. The person responsible for that line explained that this was due to a problem with the flow meter. Before the performance measures were developed, he would just have tolerated it and waited until someone from another department (responsible for the meters) made the discovery and took action to resolve it. Now, however, he took the initiative himself to have that department solve the problem quickly. Overcoming this faulty metering immediately increased the validity of the measurement data. The management also used this data for their own performance measures. Hence not only the quality of the maintenance technicians' performance measures improved, but also the quality of the performance measures of the managers.

4.4. Alternative explanations

Section 3.2.2 showed that the maintenance department faced some significant changes at the time of the development of the performance measures. These changes may have influenced the attitude of the maintenance technicians, and thus provided an alternative explanation for our findings. First of all, the company was being reorganized with the expectation of lay-offs, resulting in insecurity among the maintenance technicians. When the maintenance technicians were filling in the questionnaire, many cynically remarked that we had arrived with "perfect timing." Asking them for clarification often resulted in a response like: "Because of the current reorganization, everybody is very negative." Yet in order to avoid losing their jobs the reorganization may have triggered the maintenance technicians to work harder. While losing their jobs based on their performance was not very likely,[7] the upcoming lay-offs in the maintenance department may have given some workers a sense of urgency about the need to improve. Indeed, the next example shows that some of the registered improvements were anomalous—they could not be explained by an increase in improvement initiatives after the performance measures were developed.

[7] In accordance with Dutch labor-law regulations, the selection of which maintenance technicians were to lose their jobs was based on criteria of age and tenure (last-in, first-out per age group), rather than performance.

Example 8—rejection of empty bottles on bottling lines 2 and 4. After the performance measures were introduced three of four bottling lines showed a small but statistically significant ($p < .05$) improvement in the empty bottle rejection rate (Table 12). Yet the action researcher who often attended daily line meetings never observed any discussions between the maintenance technicians about this performance aspect, nor any overt attempt to improve the reported performance. So besides a possible contagion effect, there was no evidence whatsoever the developed performance measures had anything to do with that improvement. Hence there may have been another force—such as the reorganization—that caused this effect. However, the performance improvement in the other examples—that supported our propositions—is much higher than the improvement shown in Example 8. In other words, the best inference to draw is that the improvement initiatives after the performance measures were developed probably had an incremental effect on the performance, more than any other factor. Therefore, in general it is reasonable to contend that the employees' involvement in and influence on the development of the performance measures played a key part in the realized improvements.

Another important change was the recent replacement of one of the two maintenance managers, as mentioned in Section 3. While it is difficult to compare their performance because the lines they supervised were so different, we did see performance improvements in the lines of both managers. Thus, it does not seem likely that differences between these managers provide alternative explanations for the reported results.

Finally, our entire "package" of the intervention to develop performance measures in a participatory way will have contributed to an increase in *employee initiative*, rather than only "participation." For example, the simple fact that the employees were told from the beginning they would be expected to take more initiative may have explained the increase in initiative. However, this was an important step in the intervention process and it is consistent with TPB being transparent and explicit about the intended behavioral change of participants. This entire project was not an experiment wherein the objectives should be kept secret from the subjects. To the contrary, we think telling the objective was an important element of the approach taken for participatively developing performance measures—albeit not sufficient. Perhaps the social pressure towards taking initiative increased a bit, because it made the technicians start to recognize what was expected of them. But it is unlikely that it would have an influence on *attitude* and *capability*. Since *capability* seems to be a necessary condition to increase employee initiative (see Section 4.1.3), we think more was needed than just communicating the purpose of the project.

Another possible alternative explanation for the increase in *employee initiative* with regard to the development process that was used is the fact the process started with thinking of improvement ideas before the performance measures were even developed. Again, this was helpful for reaching the goals and a deliberate part of the participative approach for developing performance measures, but not sufficient. We only saw attempts to actually improve after the performance measures were in use, but not immediately after the brain-write sessions in which the technicians had to write down as many improvement ideas as possible. If those early meetings in September 2008 had indeed led to more *employee initiative*, we would have detected improvements in departmental performance by October or at least November. However, the evidence in Tables 4–7 and 11 tells a different story—the first improvements were realized only right after the measures were put to use.

5. Discussion

In this study, we developed a model that explains why *PM participation* influences *employee initiative*. We provided empirical support for the propositions. Our main findings showed that the performance measures developed in a participatory fashion can improve: (1) *attitude*—due to feedback on the outcomes of improvement initiatives; (2) *social pressure*—because it provided the maintenance technicians with shared priorities and targets; and (3) *capability*—because the performance measures uncovered various improvement opportunities. These variables in turn positively influenced *employee initiative*. Questionnaire results show that all three—*attitude*, *social pressure*, and *capability*—are significantly correlated with *employee initiative*. However, only the relation with *capability* remains significant when all the variables are analyzed at the same time.

We found no support for alternative explanations, and we found one unexpected strong relation, namely a correlation between *attitude* and *social pressure* to take initiative. This supports a slightly different representation of our model wherein *social pressure* indirectly leads to *employee initiative* via *attitude* to take initiative (cf. Chang, 1998; Vallerand et al., 1992). Chang's (1998) explanation for this is that people base their attitude towards performing a certain behavior on how others who are important to them consider the behavior. Our qualitative results provide some suggestive support for this interpretation: in the examples we saw that the *attitude* to take initiative was mainly influenced by the feedback the employees received from the performance measures that they had developed together with peers and managers, which are both manifestations of *social pressure*.

Describing how operational employees were involved in the process of developing performance measures is a further contribution of this research, because this bottom–up approach has received little attention in the accounting literature so far (Otley, 1999; Abernethy and Bouwens, 2005). Importantly, we made it clear from the beginning that the performance measures were intended to help the employees taking the initiative to improve the performance of their department, and not as a control device for management. To make sure that the performance measures were in line with the goals of the organization, the technicians were divided into four groups. The process began by soliciting operational improvement ideas during the groups' meetings, using a so-called brain-write. Performance measures were then developed iteratively at several subsequent group sessions. In many of these meetings

Table 12
Results of "rejection of empty bottles" (% in relation to production).

Responsible manager	Production line	Months before measure was finished							Months after measure was finished				
		Jun-08	Jul-08	Aug-08	Sep-08	Oct-08	Nov-08	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09
Returnable	2a[a,b]	1.2%	1.1%	1.3%	1.9%	1.7%	1.1%	1.1%	0.8%	0.7%	–	1.0%	1.1%
	2b[a,b]	2.5%	2.6%	1.5%	1.7%	3.5%	1.4%	1.4%	1.2%	1.2%	1.5%	1.4%	–
	3	1.0%	0.8%	0.8%	1.4%	1.6%	1.4%	1.1%	1.0%	0.8%	1.1%	1.0%	1.2%
	4[a]	3.2%	3.3%	3.7%	3.6%	3.1%	2.6%	3.4%	2.5%	2.9%	2.7%	2.8%	2.0%

[a] The T-test shows a significant improvement since implementation ($p < .05$): line 2a: $T(8) = 2.76$; line 2b: $T(5) = -2.40$; line 4: $T(10) = 2.10$.
[b] Line 2 uses two kinds of bottles (indicated by "2a" and "2b") that differ a lot on this measure. The maintenance technicians found it more useful to measure them separately.

prototype versions which were based on actual data were discussed (Wouters and Roijmans, 2011). The process was facilitated in a *nuanced* way. The main action researcher presented herself as a process facilitator who would help the employees to get their own ideas to work and thus increase productivity, instead of as an expert who introduces contextually ambiguous new ideas. She sought to maintain a careful balance between listening and proposing new measurement ideas. She had a broad knowledge of the performance measurement literature and previous performance measurement projects, and she was familiar with complex information systems. She used this expertise to not only assure their engagement for this work; she had a far more demanding job—asking countless questions and follow-ups; building collaborative prototypes; asking for continual feedback and resolutions; bringing fresh ideas to the table; and challenging constructively extant ideas, etc.

Since the action researcher plays a key role within the process of developing the performance measures, a relevant question is if the results are driven by the researcher instead of the development process: would the results have been the same had another action researcher directed the actions, or would the same researcher have achieved the same results in other ways? In Section 3.3 we tried to be very clear about the intervention in order to make it replicable. In fact, a very similar intervention has been conducted among the employees of a public sector call center by another action researcher (Gravesteijn et al., 2011; Groen et al., 2011). In that study it was found that employees also showed more *employee initiative*, resulting in many small performance improvements. In both projects the facilitative project-management role of the action researcher as well as the new participatively built performance measures seemed essential. We cannot conclude definitively whether the same researcher would have achieved the same results in other ways, but we do believe that such would be very unlikely.

Developing performance measures together with the maintenance technicians had a positive effect on their *attitude*, *social pressure*, and *capability* to take initiative, which in turn affected their behavior regarding taking more initiatives for performance improvement. To affect behavior on a continuing basis, *attitude*, *social pressure* and *capability* should be kept at the same level as after the intervention, until the new behavior becomes habitual (Ajzen, 1991). Our model does not extend to that longer-term aim. We only explain and observe behavior in direct relationship to the intervention aimed at changing the behavior in the near term; sustaining the desired behavior is another critically important issue but is not within the scope of this study.

Since we found a positive effect of an intervention on the behavior of employees, a comparison with the Hawthorne studies is relevant. These studies showed a change in employee behavior after the employees participated in an intervention that could not be explained by the intervention itself. This is often termed "the Hawthorne effect." In hindsight the behavioral changes in these classical studies were explained in several ways, such as due to changes in employees' attitude, interpersonal relationships, acquiring skill, awareness of being under study, continuous feedback, or supervision (Wickström and Bendix, 2000). We explicitly addressed similar effects in the present study. The first three alternative explanations for the Hawthorne studies' results are included in our model in the form of *attitude*, *social pressure*, and *capability*. We do not know whether awareness of being under study played a part in the results, but we do know that influence of continuous feedback and supervision was present in this study. These were part of our intervention and necessary to develop useful performance measures together with the employees, and to eventually get the positive changes in behavior. However, as similarly discussed in Section 4.4, just conditions of being under study and continuous feedback and supervision do not explain why improvements were only found immediately after the performance measures were in use. This supports our conclusion that participatorily-developed performance measures may positively affect *employee initiative*.

Limitations of our research design are that the results are built on only one company, and that we do not know if all the relations hold were they analyzed together in one model. In addition, since we only developed the performance measures in a participative way, it was not possible to compare it to a situation in which performance measures were made without the participation of employees. It would be desirable to conduct a large-scale, cross-sectional quantitative study, testing the whole model with varying degrees of participation. Furthermore, inasmuch as action research is inherently an iterative and selective process of theory development and data gathering, researcher bias may play a role (Maxwell, 2005).

Given these caveats, the fact remains that the strength of this research method is that it allowed the gathering of triangulated data, including the observing of the processes first-hand. From the start, we were challenged to demonstrate that company-university cooperation could lead to innovative results that could be implemented straightaway and be of practical relevance to the company. The employees were surprisingly cooperative and helpful in trying to make their work more measureable. There was a remarkable change from "this won't work in our situation" to "now we know what performance measures can do for us." We found that positive effects were brought about despite—or maybe because of—the fact that performance measures were not used for formal evaluations by management. The employees became quickly engaged and expected that spending time with the researchers would be worthwhile *for them*. It was extraordinarily interactive, the complete opposite from the commonplace top–down linear process where the researchers design frameworks and the company implements them. Our journey of collaborative discovery (Van de Ven and Johnson, 2006) helped to better understand how employees can together develop their own departmental performance measures, and why this may lead them to take useful initiatives for operational performance improvement.

Role of the funding source

The company of our study provided financial support for the research, in exchange for which the action researcher spent 60% of her time at the company to develop and implement performance measures together with the departmental employees. The goals formulated by the company, following several discussions with the researchers, were (1) stimulating employees to take more initiative and (2) increasing the performance of the department. Consequently the study was designed as action research, to be focused on *employee initiative* and *departmental performance*. We were free to use any other instrument that would help us to reach our practical and/or scientific goals. Data collection and analysis were done by one action researcher who was guided in the entire process by two university-based senior researchers. The report was written by the three researchers, and after completion the company consented to its publication under the condition of making minor adjustments only in terms of the level of detail in which the departmental performance data were presented in Table 4, because it revealed production volume data.

Acknowledgements

The authors thank the following persons for their helpful comments on earlier versions of this paper: the journal's reviewers; Sally Widener; Sander van Triest; Matthew Hall; Arnold Ross; Jeff Hicks; Jadzia Siemienski-Kleyn; Paul Bakker and Nóra Szűcs; as well as conference reviewers for and session participants at the Fifth Conference on Performance Measurement and Management Control, the American Accounting Association 2010 Management Accounting Section Meeting, the European Accounting Association Annual Congress 2010, the European Academy of Management Conference 2011, the Academy of Management 2011 Annual Meeting, the University of Hong Kong, and the University of Twente. A previous version of this paper has received the 2011 Best Action Research Paper Award of the ODC Division of the Academy of Management. We also thank the company for participating in this research, also financially, and the employees for their cooperation and collegiality.

Appendix A. Measurement instruments

Answering format for all items in the questionnaire: 1. Totally disagree–7. Totally agree.

A.1. Items "attitude to take initiative"

1. Thinking up and carrying out improvement ideas by myself is pleasant.
2. Thinking up and carrying out improvement ideas by myself is useful.
3. Thinking up and carrying out improvement ideas by myself is positive.
4. Thinking up and carrying out improvement ideas by myself is good.

A.2. Items "social pressure to take initiative"

Most people within «the company» who are important to me. . .

1. . . . expect of me to think up and carry out improvement ideas by myself.
2. . . . want me to think up and carry out improvement ideas by myself.
3. . . . think that I should think up and carry out improvement ideas by myself.
4. I feel social pressure to think up and carry out improvement ideas by myself.

A.3. Items "capability to take initiative"

1. I am confident that I could think up and carry out improvement ideas by myself.
2. It is easy for me to think up and carry out improvement ideas by myself.
3. There are factors that make it difficult for me to think up and carry out improvement ideas by myself (recoded and deleted).
4. It is possible for me to think up and carry out improvement ideas by myself (deleted).

A.4. Items "employee initiative"

1. I actively attack problems.
2. Whenever something goes wrong, I search for a solution immediately.
3. Whenever there is a chance to get actively involved, I take it.
4. I take initiative immediately even when others don't.
5. I use opportunities quickly in order to attain my goals.
6. Usually I do more than I am asked to do.
7. I am particularly good at realizing ideas.

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. Administrative Science Quarterly 41, 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. Contemporary Accounting Research 21 (2), 271–301.

Ajzen, I., 1991. The theory of planned behavior. Organizational Behavior and Human Decision Processes 50 (2), 179–211.

Ajzen, I., 2006. Behavioral interventions based on the theory of planned behavior. Retrieved June 26, 2008 from http://people.umass.edu/aizen/pdf/tpb.intervention.pdf.

Ajzen, I., 2006. Constructing a TPB questionnaire: conceptual and methodological considerations. Retrieved July 10, 2008 from http://people.umass.edu/aizen/pdf/tpb. measurement.pdf.

Anderson, N., Herriot, P., Hodgkinson, G.P., 2001. The practitioner-researcher divide in industrial work and organizational (IWO) psychology: where are we now, and where do we go from here? Journal of Occupational and Organizational Psychology 74, 391–411.

Axtell, C.M., Parker, S.K., 2003. Promoting role breadth self-efficacy through involvement, work design, and training. Human Relations 56, 112–131.

Baskerville, R., Wood-Harper, A.T., 1998. Diversity in information systems action research methods. European Journal of Information Systems 7, 90–107.

Becker, T.E., 2005. Potential problems in the statistical control of variables in organizational research: a qualitative analysis with recommendations. Organizational Research Methods 8 (3), 274–289.

Birnberg, J.G., Luft, J.L., Shields, M.D., 2007. Psychology theory in management accounting research. In: Chapman, C.S., Hopwood, A.G., Shields, M.D. (Eds.), Handbook of Management Accounting Research (Vol. 1). Elsevier, Oxford, UK, pp. 113–136.

Bisbe, J., Otley, D., 2004. The effects of the interactive use of management control systems on product innovation. Accounting, Organizations and Society 29, 709.

Bono, J.E., Judge, T.A., 2003. Self-concordance at work: toward understanding the motivational effects of transformational leaders. Academy of Management Journal 48 (5), 554–571.

Bourne, M., 2005. Researching performance measurement system implementation: the dynamics of success and failure. Production Planning & Control 16 (2), 101–113.

Bourne, M., Neely, A., Platts, K., Mills, J., 2002. The success and failure of performance measurement initiatives. International Journal of Operations and Production Management 22 (11), 1288–1310.

Brehm, S.S., Kassin, S.M., Fein, S., 2002. Social Psychology, Fifth ed. Houghton Mifflin Company, Boston, MA.

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. Accounting, Organizations and Society 34, 305–321.

Burney, L.L., Widener, S.K., 2007. Strategic performance measurement systems job-relevant information, and managerial behavioral responses: role stress and performance. Behavioral Research in Accounting 19, 43–69.

Campbell, D.J., 2000. The proactive employee: managing workplace initiative. Academy of Management Executive 14 (3), 52–66.

Cavalluzzo, K.S., Ittner, C.D., 2004. Implementing performance measurement innovations: evidence from government. Accounting, Organizations and Society 29, 243–267.

Chang, M.K., 1998. Predicting unethical behavior: a comparison of the theory of reasoned action and the theory of planned behavior. Journal of Business Ethics 17, 1825–1834.

Checkland, P., Holwell, S., 2007. Action research: its nature and validity. Information Systems Action Research 13, 3–17.

Chenhall, R.H., 2005. Integrative strategic performance measurement systems, strategic alignment of manufacturing, learning and strategic outcomes: an exploratory study. Accounting, Organizations and Society 30, 395–422.

Chiles, A.M., Zorn, T.E., 1995. Empowerment in organizations: employees' perceptions of the influences of empowerment. Journal of Applied Communication Research 23, 1–25.

Collins, F., 1982. Managerial accounting systems and organizational control: a role perspective. Accounting, Organizations and Society 7 (2), 107–122.

Coughlan, P., Coghlan, D., 2002. Action research for operations management. International Journal of Operations and Production Management 22 (2), 220–240.

Covaleski, M.A., Evans, J.H., Luft, J.L., Shields, M.D., 2003. Budgetting research: three theoretical perspectives and criteria for selective integration. Journal of Management Accounting Research 15, 3–49.

Crant, J.M., 2000. Proactive behavior in organization. Journal of Management 26, 435–462.

Davis, S., Albright, T., 2004. An investigation of the effect of balanced scorecard implementation on financial performance. Management Accounting Research 15, 135–153.

De Geuser, F., Mooraj, S., Oyon, D., 2009. Does the balanced scorecard add value? Empirical evidence on its effect on performance. European Accounting Review 18, 93–122.

De Haas, M., Algera, J.A., 2002. Demonstrating the effect of the strategic dialogue: participation in designing the management control system. Management Accounting Research 13, 41–69.

De Haas, M., Kleingeld, A., 1999. Multilevel design of performance measurement systems: enhancing strategic dialogue throughout the organization. Management Accounting Research 10, 233–261.

De Leeuw, S., Van den Berg, J., 2010. Improving operational performance by influencing shopfloor behavior via performance management practices. Journal of Operations Management, doi:10.1016/j.jom.2010.12.009.

Demski, J.S., Feltham, G.A., 1976. Cost Determination: A Conceptual Approach. Iowa State University Press, Ames, IA.

Den Hartog, D.N., Belschak, F.D., 2007. Personal initiative, commitment and affect at work. Journal of Occupational and Organizational Psychology 80, 601–622.

Derfuss, K., 2009. The relationship of budgetary participation and reliance on accounting performance measures with individual-level consequent variables: a meta-analysis. European Accounting Review 18 (2), 203–239.

Dubois, A., Gadde, L.E., 2002. Systemic combining: an abductive approach to case research. Journal of Business Research 55, 553–560.

Dunn, J., Schweitzer, M.E., 2005. Why good employees make unethical decisions: the role of reward systems, organizational culture, and managerial oversight. In: Kidwell, R.E., Martin, C.L. (Eds.), Managing Organizational Deviance. Sage, Thousand Oaks, CA, pp. 39–60.

Eldenburg, L., Soderstrom, N., Willis, V., Wud, A., 2010. Behavioral changes following the collaborative development of an accounting information system. Accounting, Organizations and Society 35, 222–237.

Erez, M., Arad, R., 1986. Participative goal-setting: social, motivational, and cognitive factors. Journal of Applied Psychology 71 (4), 591–597.

Erez, M., Early, P.C., Hulin, C.L., 1985. The impact of participation on goal acceptance and performance: a two-step model. Academy of Management Journal 28 (1), 50–66.

Farrell, A.M., Kadous, K., Towry, K.L., 2008. Contracting on contemporaneous versus forward-looking measures: an experimental investigation. Contemporary Accounting Research 25 (3), 773–802.

Ferreira, A., Otley, D., 2009. The design and use of performance management systems: an extended framework for analysis. Management Accounting Research 20, 263–282.

Fishbein, M., Ajzen, I., 2010. Predicting and Changing Behavior. Psychology Press, Taylor & Francis Group, New York, NY.

Francis, J.J., Eccles, M.P., Johnston, M., Walker, A., Grimshaw, J., Foy, R., et al., 2004. Constructing Questionnaires Based on the Theory of Planned Behavior: A Manual for Health Services Researchers. Retrieved June 25, 2008, from http://www.rebeqi.org/ViewFile.aspx?itemID=212.

Franco-Santos, M., Kennerley, M., Micheli, P., Martinez, V., Mason, S., Marr, B., Gray, D., Neely, A., 2007. Towards a definition of a business performance measurement system. International Journal of Operations and Production Management 27 (8), 784–801.

Free, C., 2007. Supply-chain accounting perspectives in the UK retail sector: enabling or coercive collaboration? Contemporary Accounting Research 24 (3), 897–933.

Frese, M., Fay, D., 2001a. Personal initiative: an active performance concept for work in the 21st century. Research in Organizational Behavior 23, 133–187.

Frese, M., Fay, D., 2001b. The concept of personal initiative: an overview of validity studies. Human Performance 14 (1), 97–124.

Frese, M., Fay, D., 2003. Measuring personal initiative. Retrieved April 3, 2009, from http://www.uni-giessen.biz/content/downloads/measurement/InterviewPersInit.pdf.

Frese, M., Fay, D., Hilburger, T., Leng, K., Tag, A., 1997. The concept of personal initiative: operationalization, reliability and validity in two German samples. Journal of Occupational and Organizational Psychology 70, 139–161.

Fried, Y., Ferris, G.R., 1987. The validity of the job characteristics model: a review and a meta-analysis. Personnel Psychology 40, 287–322.

Fuller, J.B., Marler, L.E., Hester, K., 2006. Promoting felt responsibility for constructive change and proactive behavior: exploring aspects of an

elaborated model of work design. Journal of Organizational Behavior 27, 1089–1120.
Grafton, J., Lillis, A.M., Widener, S.K., 2010. The role of performance measurement and evaluation in building organizational capabilities and performance. Accounting, Organizations and Society 35, 689–706.
Gravesteijn, M.E., Evers, J.F., Molenveld, M., Wilderom, C.P.M., 2011. Leren van presteren op de werkvloer via zelfontwikkelde prestatie-indicatoren. Management en Organisatie (5), 60–79.
Groen, B.A.C., Evers, J.F., Gravesteijn, M.E., Molenveld, M., Schopman, M., Veerbeek, R., Wilderom, C.P.M., 2011. Team learning through bottom-up development of team performance indicators. Paper Presented at the 2011 Academy of Management Annual Meeting, August 12-16, 2011, San Antonio, Texas.
Hackman, J.R., Oldham, G.R., 1976. Motivation through the design of work: test of a theory. Organizational Behavior and Human Performance 16, 250–279.
Hall, M., 2008. The effect of comprehensive performance measurement systems on role clarity, psychological empowerment and managerial performance. Accounting, Organizations and Society 33, 141–163.
Hall, M., 2010. Do comprehensive performance measurement systems help or hinder managers' mental model development? Management Accounting Research, doi:10.1016/j.mar.2010.10.002.
Henri, J.-F., 2006. Management control systems and strategy: a resource-based perspective. Accounting, Organizations and Society 31, 529–558.
Hill, M., Mann, L., Wearing, A.J., 1996. The effects of attitude, subjective norm and self-efficacy on intention to benchmark: a comparison between managers with experience and no experience in benchmarking. Journal of Organizational Behavior 17, 313–327.
Hunton, J.E., Gibson, D., 1999. Soliciting user-input during the development of an accounting information system: investigating the efficacy of group discussion. Accounting, Organizations and Society 24, 597–618.
Ilgen, D.R., Fisher, C.D., Taylor, M.S., 1979. Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology 64, 349–371.
Ittner, C.D., Larcker, D.F., Randall, T., 2003. Performance implications of strategic performance measurement in financial services firms. Accounting, Organizations and Society 28, 715–741.
Jeong, S.B., 2006. Participation and Goal Setting: An Examination of the Components of Participation. Ohio State University, Ohio.
Jimmieson, N.L., Peach, M., White, K.M., 2008. Utilizing the theory of planned behavior to inform change management. Journal of Applied Behavioral Science 44 (2), 237–262.
Johns, G., Xie, J.L., Fang, Y., 1992. Mediating and moderating effects in job design. Journal of Management 18, 657–676.
Kasanen, E., Lukka, K., Siitonen, A., 1993. The constructive approach in management accounting research. Journal of Management Accounting Research 5, 243–264.
Kelly, K., 2010. Accuracy of relative weights on multiple leading performance measures: effects on managerial performance and knowledge. Contemporary Accounting Research 27 (2), 577–608.
Kleingeld, A., Van Tuijl, H., Algera, J.A., 2004. Participation in the design of performance management systems: a quasi-experimental field study. Journal of Organizational Behavior 25, 831–851.
Latham, G.P., 2003. Goal setting: a five-step approach to behavior change. Organizational Dynamics 32, 309–318.
Lee, C.-L., Yang, H.-J., 2010. Organizational structure, competition and performance measurement systems and their joint effects on performance. Management Accounting Research, doi:10.1016/j.mar.2010.10.003.
Li, P., Tang, G., 2009. Performance measurement design within its organisational context—evidence from China. Management Accounting Research 20, 193–207.
Luckett, P.F., Eggleton, I.R.C., 1991. Feedback and management accounting: a review of research into behavioural consequences. Accounting, Organizations and Society 16 (4), 371–394.
Luft, J.L., Shields, M.D., 2007. Mapping management accounting: graphics and guidelines for theory-consistent empirical research. In: Chapman, C.S., Hopwood, A.G., Shields, M.D. (Eds.), Handbook of Management Accounting Research (Vol. 1). Elsevier, Oxford, UK, pp. 27–98.
Malina, M.A., Nørreklit, H.S.O., Selto, F.H., 2007. Relations among measures, climate of control, and performance measurement models. Contemporary Accounting Research 24 (3), 935–982.
Masquefa, B., 2008. Top management adoption of a locally driven performance measurement and evaluation system: a social network perspective. Management Accounting Research 19, 182–207.
Maxwell, J.A., 2005. Qualitative Research Design: An Interactive Approach, Vol. 41., second ed. Sage, Thousand Oaks, CA.
McKinnon, S.M., Bruns Jr., W.J., 1992. The Information Mosaic. Harvard 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. Journal of Operations Management 22, 209–217.
Miller, D., Greenwood, R., Hinnings, B., 1997. Creative chaos versus munificent momentum: the schism between normative and academic views of organizational change. Journal of Management Inquiry 6, 71–78.
Moers, F., 2006. Performance measure properties and delegation. The Accounting Review 81 (4), 897–924.
Morrison, E.W., 2006. Doing the job well: an investigation of pro-social rule breaking. Journal of Management 32, 5–28.
Neely, A.D., Bourne, M., Mills, J., Platts, K., Richards, H., 2002. Getting the Measure of Your Business. University Press, Cambridge, UK.
Nørreklit, H., 2000. The balance of the balanced scorecard: a critical analysis of some of its assumptions. Management Accounting Research 11, 65–88.
Otley, D., 1999. Performance management: a framework for management control systems research. Management Accounting Research 10, 363–382.
Parker, S.K., Wall, T.D., Jackson, P.R., 1997. 'That's not my job': developing flexible employee work orientations. Academy of Management Journal 40, 899–929.
Parker, S.K., Williams, H.M., Turner, N., 2006. Modeling the antecedents of proactive behavior at work. Journal of Applied Psychology 91 (3), 636–652.
Quinn, R.E., Spreitzer, G.M., 1997. The road to empowerment: seven questions every leader should consider. Organizational Dynamics 26 (2), 37–49.
Rynes, S.L., Bartunek, J.M., Daft, R.L., 2001. Across the great divide: knowledge creation and transfer between practitioners and academics. Academy of Management Journal 44, 340–355.
Said, A.A., HassabElnaby, H.R., Wier, B., 2003. An empirical investigation of the performance consequences of nonfinancial measures. Journal of Management Accounting Research 15, 193–223.
Schein, E., 1987. The Clinical Perspective in Fieldwork. Sage, Newbury Park, NJ.
Sheldon, K.M., Elliot, A.J., 1998. Not all personal goals are personal: comparing autonomous and controlled reasons for goals as predictors of effort and attainment. Personality and Social Psychology Bulletin 24 (5), 546–557.
Shields, J., Shields, M.D., 1998. Antecedents of participative budgeting. Accounting, Organizations and Society 23 (1), 49–76.
Spreitzer, G.M., 1995. Psychological empowerment in the workplace: dimensions, measurement, and validation. Academy of Management Journal 38, 1442–1465.
Spreitzer, G.M., 1996. Social structural characteristics of psychological empowerment. Academy of Management Journal 39, 483–504.
Sprinkle, G.B., 2003. Perspectives on experimental research in managerial accounting. Accounting, Organizations and Society 28, 287–318.
Strauss, A., Corbin, J., 1990. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage, London.
Terhürne, H., 2008. Breinschrijven. Management Tools 4, 22–23.
Thomas, K.W., Velthouse, B.A., 1990. Cognitive elements of empowerment: an interpretive model of intrinsic task motivation. Academy of Management Review 15, 666–681.
Thompson, L., 2003. Improving the creativity of organizational work groups. Academy of Management Executive 17 (1), 96–111.
Vallerand, R.J., Deshaies, J., Cuerrier, J.-P., Pelletier, L.G., Mongeau, C., 1992. Ajzen and Fishbein's theory of reasoned action as applied to moral behavior: a confirmatory analysis. Journal of Personality and Social Psychology 62, 98–109.
Van de Ven, A., Johnson, P.E., 2006. Knowledge for theory and practice. Academy of Management Review 31 (4), 802–821.
Van Veen-Dirks, P., 2009. Different uses of performance measures: the evaluation versus reward of production managers. Accounting, Organizations and Society 35, 141–164.
Vroom, V.H., 1995. Work and Motivation, First ed. Jossey-Bass, San Francisco, CA.
Webb, R.A., 2004. Managers' commitment to the goals contained in a strategic performance measurement system. Contemporary Accounting Research 21 (4), 925–958.
Widener, S., 2006. Associations between strategic resource importance and performance measure use: the impact on firm performance. Management Accounting Research 17, 433–457.
Widener, S., 2007. An empirical analysis of the levers of control framework. Accounting, Organizations and Society 32, 757–788.

Wickström, G., Bendix, T., 2000. The Hawthorne effect: what did the original Hawthorne studies actually show? Scandinavian Journal of Work, Environment & Health 26 (4), 363–367.

Wilderom, C.P.M., Wouters, M.J.F., Van Brussel, J., 2007. Attitudes toward developmental performance measurement: professionalism, team trust and leadership. Paper Presented at the 2007 Annual meeting of the Academy of Management. Philadelphia.

Wouters, M.J.F., 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.J.F., Wilderom, C.P.M., 2008. Developing performance-measurement systems as enabling formalization: a longitudinal field study of a logistics department. Accounting, Organizations and Society 33, 488–516.

Available online at www.sciencedirect.com

ScienceDirect

Accounting, Organizations and Society 33 (2008) 488–516

Accounting, Organizations and Society

www.elsevier.com/locate/aos

Developing performance-measurement systems as enabling formalization: A longitudinal field study of a logistics department

Marc Wouters *, Celeste Wilderom

University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Abstract

This paper reports on a developmental approach to performance-measurement systems (PMS). In particular, we look at characteristics of a development process that result in the PMS being perceived by employees as enabling of their work, rather than as primarily a control device for use by senior management. We will refer to such a PMS as "enabling PMS". The theoretical part of the study builds on ideas of enabling versus coercive formalization [Adler, P. S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly 41* (March), 61–89]; on notions of organizational learning (e.g., [Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science 13*(3), 339–351]); and on awareness of the incompleteness of performance measures (e.g., [Chapman, C. S. (1997). Reflections on a contingent view of accounting. *Accounting, Organizations and Society 22*, 189–205; Lillis, A. M. (2002). Managing multiple dimensions of manufacturing performance—An exploratory study. *Accounting, Organizations and Society 27*, 497–529]). The empirical context entails a mixed-method, 3-year longitudinal study of the logistics department of a medium-sized company in the beverage manufacturing industry. Qualitative data were gathered through interviews, participation in meetings, action research, and review of company documents. We also analyzed two waves of quantitative survey data, gathered from a panel of 42 employees. We find that a development process that is *experience-based* contributes to the enabling nature of the PMS, as it builds on existing skills, local practices, and know-how on performance measurement to enrich the PMS step-by-step over time. Also, *experimentation* with specific performance measures was found to enhance the enabling nature of the PMS: testing, reviewing, and refinement of conceptualizations, definitions, data, and presentations of new performance measures. *Professionalism* was significantly related to positive attitude toward performance measures in our survey data. The results also illustrate that *transparency* of the PMS itself is key to enabling PMS.

* Corresponding author. Tel.: +31 53 4894498; fax: +31 53 4892159.
E-mail address: m.j.f.wouters@utwente.nl (M. Wouters).

0361-3682/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aos.2007.05.002

Introduction

Performance-measurement systems (PMS) are mostly studied from the perspective of top-management: how it allows them to monitor whether given objectives have been achieved. In addition, it may also help top managers to formulate strategy, specify operational actions needed for implementation, set targets in relation to current performance (so as to reveal priorities for operational improvement), and clarify mutual expectations (Abernathy & Brownell, 1999; Bisbe & Otley, 2004; de Haas & Algera, 2002; Simons, 1990, 1991, 1994, 1995). However, what about the managers who are the subject of PMS—whose performance is being measured? There are few accounts in the PMS literature where lower and middle-level employees and managers consider a PMS as something that supports *them*, that they can use for their own purposes to assess how things are going, identify problems, prioritize issues, develop ideas for improvement, engineer solutions for concrete problems, or make decisions (Jönsson & Grönlund, 1988). We refer to "enabling PMS" when it is perceived by employees as enabling of their work, rather than as primarily a control device for use by senior management.

This study investigates performance-measurement systems in operations, closely connected to the specifics of particular operational processes. Building on Adler and Borys (1996), we conceive of a PMS as a form of formalization. *Coercive formalization* aims to force employee compliance, while *enabling formalization* makes employees feel facilitated or motivated by the rules and the systems in place. Adler and Borys (1996) contrast enabling and coercive types of formalization along three dimensions: (1) characteristics of the system, (2) the process of designing the system, and (3) the implementation of the system. These dimensions are relevant for understanding the role of management control in organizations, as demonstrated by Ahrens and Chapman (2004). While their study primarily focused on (1) characteristics of the system—in terms of repair, transparency and flexibility—we focus on points (2) and (3): the *development process* for designing and implementing the performance-measurement system. We expect that the manner by which the development process is carried out affects the extent to which the PMS will be perceived by employees as enabling. In this paper we address the question: Which characteristics of a PMS development process enhance the enabling nature of the PMS?

Previous research has shown that developing an enabling PMS is a delicate process. Townley, Cooper, and Oakes (2003), for example, demonstrated that while the introduction of performance measures may begin as an initiative considered by various levels in the organization to be nuanced, supporting and constructive, the process may easily derail: "From an initial discourse that emphasized a potential for reasoned justification, debate and dialogue quickly collapsed into a standard template" (Townley et al., 2003, p. 1058). Qu (2006) found that consultants considered the incorporation of client input—and especially information on existing reports and specific measures already in use—crucial for the production of a usable PMS. Failure to include such input was a major source of frustration for participants in the development process (Qu, 2006). Yet, there is little empirical knowledge about what kind of a development process fosters an enabling PMS. "The balanced scorecard literature also indicates that it [is] as much the *process* of establishing a scorecard that yields benefit as the resultant measurement schema. However, the literature is remarkably silent on this point" (Otley, 1999, p. 377, emphasis added).

This study aims to contribute to the literature by theoretically and empirically investigating characteristics of a PMS development process that enhance the enabling nature of the PMS. We build on the framework of Adler and Borys (1996), who propose that user involvement and professionalism contribute to enabling formalization. We develop these ideas further in the context of PMS. We consider the inherent incompleteness of PMS in terms of the inability to reflect the various dimensions of operational performance and tradeoffs among these (Lillis, 2002), and therefore user involvement needs to be mobilized, both in terms of existing experience with quantification of performance, and also throughout the design and implementation process of new measures. Design and implementation include several activities, such as shaping and further improving the best fitting definitions of useful performance measures; finding or creating measurement data for determining the actual values of these performance measures; building information systems for reporting performance-measurement results; setting performance level targets for performance measures; and periodically reviewing, revising and refining both single measures and the overall PMS. We look at such activities from the perspective of how organizations can learn by carefully building on and reusing existing experiences (cf. Zollo & Winter, 2002), and experimenting and prototyping with new practices (cf. Carlile, 2002). ‘‘Design'' and ‘‘implementation'' are hard to distinguish (Adler & Borys, 1996), and we prefer to combine them and use the phrase ‘‘development process''. This reflects that design and implementation activities are conducted in a mutually constitutive, iterative fashion: employees learn through implementation, on the basis of which they adjust the design of the PMS, which leads to new implementation activities, etc. Such an approach assumes a considerable level of professionalism.

The empirical findings are based on a 3-year longitudinal case study within the logistics department of a medium-sized company in the beverage manufacturing industry. We gathered both survey and qualitative data. The research includes not only observation of the company's activities, but also elements of action research, since we were involved in the development of a departmental PMS.

The structure of this paper is as follows. Performance measures as either coercive or enabling formalization are introduced in performance measures as enabling or coercive formalization section. Our propositions regarding the characteristics of a developmental process that contributes to an enabling PMS are put forward in propositions about a developmental approach for enabling PMS section. The research methods are described in research method section. Empirical results are presented and discussed in results: developing an enabling PMS and discussion section, and conclusion is given in the final section.

Performance measures as enabling or coercive formalization

Traditionally, performance measures in operations put a one-sided emphasis on minimizing direct costs through low material costs, high capacity utilization, and high direct labor efficiency. However, early research identified the need to broaden performance-measurement systems to support new operations practices, and advocated the use of measures for quality, throughput times, flexibility, etc. (Beamon, 1999; Eccles, 1991; Hall, Johnson, & Turney, 1990; Kaplan, 1983, 1990; Maskell, 1991; Nanni, Dixon, & Vollmann, 1992). Empirical studies have supported relationships between the pursuit of specific operational strategies and the expansion of traditional efficiency-focused PMS to include new performance measures (e.g., Abernathy & Lillis, 1995; Baines & Langfield-Smith, 2003; Banker, Potter, & Schroeder, 1993; Fullerton & McWatters, 2002; Perera, Harrison, & Poole, 1997; Maiga & Jacobs, 2005).

But despite to broadening of PMS—in both research and practice—to embrace a wider portfolio of measures, the approach to developing the PMS has received far less attention in empirical studies. In this section, we will first discuss issues in regard to the incompleteness of performance-measurement systems, and thereafter we will introduce more specifically the control ideas laid out by Adler and Borys (1996).

Incompleteness of a PMS

Incompleteness of PMS arises when strategic performance measures are disaggregated into different performance dimensions, separate periods and organizational sub-units, and the dependencies between disaggregated measures are not reflected in the PMS (Lillis, 2002). For example, attempts to improve responsiveness may lead to more frequent changeovers, demands for shorter lead times, and higher inventories, and when such tradeoffs are inadequately reflected in the PMS there is a likely ‘‘friction created by the failure to determine and adjust for the implications of profit centre strategy on the manufacturing cost function'' (Lillis, 2002, p. 510). Designing a perfectly complete PMS remains challenging, if not impossible, and would require nothing less than the expression of all relevant aspects of performance in quantitative terms (financial and non-financial), estimation of the tradeoffs among such dimensions of performance in the setting of targets for financial and non-financial performance measures, and consideration of interdependencies between different organizational units (and different time periods) in the PMS (see, e.g., Lillis, 2002).

And the greater the incompleteness, the more the PMS may be perceived by functional sub-units as a ''negative'', ''unfair'', ''threatening'', or ''coercive'' instrument of management control. Malina and Selto (2001) found that perceptions of PMS were more negative if measures were inaccurate or subjective, and if benchmarks were considered inappropriate but nevertheless used for evaluation. In other words, employees may feel that their performance ''as measured'' (by the metrics) does not truthfully reflect what they see as their ''real'' contribution to the organization. For example, they may find it unfair that contingencies (uncontrollable circumstances), materializing after targets have been set, are not considered for adjusting those targets; employees may not believe their supervisors use the PMS in a fair way for evaluating their performance; employees may regard target levels as overly ambitious and unrealistic; or they may feel their personal risk has increased too much because of consequences that are tied to PMS results.

Several studies have found evidence of the relationship between the use of controls and defensive behavior—such as negotiating targets towards more easily achievable levels, obtaining surplus resources for completing tasks, concealing windfalls that have made tasks easier than anticipated, or even taking operational decisions just to make the results ''as measured'' look good at the expense of negative long-term effects—sometimes moderated by variables such as measurability of outputs, the extent to which input–output relationships of processes are understood, and the style in which the controls are used (e.g., Carmona & Grönlund, 2003; Chow, Kato, & Merchant, 1996; Jaworski & Young, 1992; Ramaswami, 1996, 2002; Van der Stede, 2000).

Several studies have identified ways in which firms manage incompleteness of PMS. Lillis (2002) found that firms sometimes loosened control reactions to variances, implemented more innovative PMS, integrated the PMS with other management systems, or used measurement weightings. Davila and Wouters (2005) described a firm that designed a budgeting system that reduced emphasis on cost targets and provided budgetary slack when performance attributes other than costs required attention. Van der Stede (2000) found that firms balanced the strictness of controls with a business unit's strategy. Business units following a differentiation strategy implemented less rigid budgetary control, which allowed for some budgetary slack and stimulated managers to think long term.

Enabling formalization

Incompleteness motivates why designing and implementing PMS in operations is difficult and requires a deliberate and careful approach. For developing our propositions about a development process that is likely to enhance the enabling nature of the PMS, we build on the framework of Adler and Borys (1996). First, because this framework conceptualizes the issue that is central to our accounting study: the distinction between performance-measurement systems that only serve higher-management needs and control employees' behavior (coercive formalization), versus systems that support employees to do their work better (by providing feedback, identifying problems, revealing improvement opportunities, help prioritizing action, etc.): enabling formalization. Second, because this framework helps to articulate that characteristics of the *system* itself, as well as *processes* for design and implementation of the system may contribute to the coercive or enabling nature of formalization. Third, because Adler and Borys (1996) offer initial suggestions about what kind of design and implementation process is likely to foster the enabling nature of formalization, and so it helps to delineate our intended contribution: to draw on organizational literature as well as the empirical material to further develop the understanding of enabling PMS development.

Adler and Borys (1996, p. 66) propose that ''employees' attitudes to formalization depend on the type of formalization with which they are confronted''. They suggest that employee attitudes are more positive when formalization enables them to better master their tasks, and will be more negative when it ''functions as a means by which management attempts to coerce employees' effort and compliance''. Enabling formalization mobilizes rather than replaces employees' intelligence, and acts to ''help users form a mental model of the system they are using'' (p. 70). As such, these kinds of ''procedures provide organizational memory that captures lessons learned from experience'' (p. 69).

It is thus relevant to better understand how organizations may achieve enabling formalization. Adler and Borys (1996) suggest that whether formalization has an enabling or coercive character, depends on characteristics of the *formalization* as well as on the *process of designing and implementing the system*. These characteristics of formalization are internal and global transparency, and flexibility and repair. We will discuss these first before we begin our analysis of a development process conducive to enabling formalization. *Internal transparency* means that users have a good understanding of the logic of a system's internal function and they have information on its status. Enabling formalization provides users with a clear understanding of the underlying rationale for why certain control mechanisms are in place. Such formalization also codifies best-practice experiences, and users are provided with feedback on their performance. *Global transparency* refers to the intelligibility for employees of the broader system and context within which they do their work. Controls are designed to afford employees an understanding of where their own tasks fit into the whole. Information from beyond one's specific domain is available. *Flexibility* means that users can make controlling decisions after enabling systems have provided information. ''Flexible systems encourage users to modify the interface and add functionality to suit their specific work demands'' (p. 74). *Repair* means that users can mend and improve the work process themselves rather than allowing breakdowns and other non-programmable events to force the work processes to a halt. We refer to Ahrens and Chapman (2004) who discuss these characteristics of enabling formalization in the context of management control systems.

As mentioned above, Adler and Borys (1996) also contrast enabling and coercive types of formalization in terms of the *processes of designing and implementing the system*. They discuss some of the characteristics of these processes that are likely to lead to enabling formalization, such as employee voice, employee skills, process control, and flexibility in changing controls. They propose that ''employee involvement in the formulation of procedures is likely to have a positive effect on both attitudinal and technical outcomes'' (p. 75). Principles for the design of equipment technology, they suggest—such as a focus on users and usability, early and continual user testing, and iterative design processes—carry over to the development of formalization as ''organizational technology''. However, the design and implementation of formalization are typically intertwined: while equipment may be bought ''off the shelf'', customized from existing modules, or designed-to-specification outside the client organization, ''organizational technology'' takes shape within the specific implementation context. Adler and Borys (1996) call for more research to explore whether and how organizations can introduce enabling types of formalization. We will build on their framework and develop their ideas further specifically in the context of PMS development.

Propositions about a developmental approach for enabling PMS

This section sets out three propositions in response to our research question: ''What characteristics of a PMS development process enhance the enabling nature of the PMS?'' We propose that a development process that is characterized by (1) being experienced-based, (2) allowing experimentation, and (3) building on employees' professionalism is likely to result in an enabling PMS. *Experienced-based* involves the identification, appreciation, documentation, evaluation, and consolidation of existing local knowledge and experience with respect to quantitatively capturing and reporting relevant aspects of performance. *Experimentation* involves the first development of a performance measure and the subsequent testing and refinement (in several rounds) of its conceptualization, definition, required data, IT tools, and presentation, together with employees (whose performance is going to be measured), to arrive at a measure that is a valid, reliable, and understandable indicator of performance in a specific local context. *Professionalism* of employees denotes an orientation toward learning for the purpose of improving work practices. We underpin these propositions in the remainder of this section; and in the section with empirical results we will discuss and illustrate them further.

We feel that presenting propositions before the empirical study helps to better discuss our theoretical ideas in relation to the literature, and empirical findings in relation to the theory; it is not to suggest that theory and findings have been developed subsequently. Rather, the nature of the research process was as discussed by Ahrens and Chapman (2006, p. 836): "Problem, theory, and data influence each other throughout the research process. The process is one of iteratively seeking to generate a plausible fit between problem, theory, and data". Before the study started, we explicitly intended to explore an experience-based development process and what we than called continuous revision of the PMS (later formulated as "experimentation"). These ideas took further shape during the course of the study through going back-and-forth between the fieldwork and the literature. Furthermore, the development of the survey instrument, which started about 15 months into the study, involved an extensive process of focusing and making connections between the field and the literature, and in this stage the role of "professionalism" was highlighted and subsequently focused upon in the fieldwork. Later in the research project, we became familiar with the framework of Adler and Borys (1996), and this was found to be a very powerful way for organizing the theoretical discussion and empirical results.

Experience-based development process

Organizational change processes may take advantage of local knowledge, which can be defined as "the very mundane, yet expert understanding of and practical reasoning about local conditions derived from lived experience" (Yanow, 2004, p. S12). Organizational change processes that utilize local knowledge are more likely to lead to sustainable changes and improvements (Abrahamson, 2000; Lowe & Jones, 2004; Zollo & Winter, 2002). In the context of PMS, we propose that a development process that is experience-based is likely to have a positive effect on the enabling nature of the PMS. An experience-based development process involves the identification, appreciation, documentation, evaluation, and consolidation of existing local knowledge and experience with respect to quantitatively capturing and reporting relevant aspects of performance. We will elaborate on the idea of an experienced-based development process in this section.

Many of the proposed approaches for design and implementation in the literature, however, seem to pay little attention to either experience or user involvement. Most approaches to PMS design and implementation (see Bourne, Neely, Mills, & Platts (2003a) for a review of the PMS development processes literature) focus on how the goals set at the top of the organization can better guide actions taken lower in the organization. First steps in the typical development process are to clearly define the overall (i.e., corporate-level) strategic objectives and then the local operations' specific contribution toward achieving these overall strategic objectives. Thus, the organization's global performance measures and functional measures are derived. The PMS is typically designed from the perspective of top-management, as is apparent in the following representative characteristics: (1) explicit reflection of the firm's strategic objectives and subsequent break-down of those objectives to more specific objectives at lower managerial levels, (2) the signaling of performance levels that are below targets, (3) the ability to "drill-down" and get more details when needed, (4) striving for transparency, consistency, and uniformity regarding definitions of performance measures, presentation formats, etc., and (5) one information system that contains all data and reports. External experts may be involved, who often bring in a standardized way of designing and implementing the system, with examples (or templates), complete with performance measures, presentation formats, and a set consulting approach for designing the system, software tools, etc.

However, top-down, mandated performance-measurement initiatives are less likely to be successful (Cavalluzzo & Ittner, 2004; Scott & Tiessen, 1999; de Haas & Algera, 2002). These well-intentioned, standardized methods carry the danger of insufficiently reflecting the local organizational contexts or the available experience and unique expertise of employees. Furthermore, even before such measurements systems are initiated, a number of informal performance measures, at various levels within the organization are already in use by managers, complementing the information they get from other sources, such as observations, or conversations with people individually or in group meetings, as well as non-face-to-face communication through phone calls or emails (McKinnon & Bruns, 1992). These informal measurement reports are often developed locally, contain a mix of local and centralized data, report operating information over a very short period of time (weeks, days, or less), provide status information (up-to-date accumulations of bits of operating data, e.g., inventory-level reports and backlog reports), and enable performance comparisons between, for example, budgeted vs. actual performance, one time period vs. another, etc. (McKinnon & Bruns, 1992). Such informal reports use a variety of presentation formats, performance measure definitions, data, and information systems. The existence of such reports is often unknown outside the organizational unit where they are produced and used, to the extent that, from the perspective of top-management, a coherent PMS does not appear to exist at all within the organization! Although employees may have considerable experience with performance measures, and may have already established context-specific practices, from the perspective of top-management these do not constitute a PMS.

Typically, expert-led approaches initiated by top-management, are not likely to expend the effort necessary to build an in-depth understanding of locally-developed existing reporting practices, in particular about the detailed definition, data, motivation for, and experiences with existing measures and information systems (Qu, 2006). The consultants are also more likely to address problems from the perspective of top-management (or whoever hires them), and they may seek to focus on concepts that are fashionable in the business literature, and to attempt to transfer their earlier experience to the project at hand (see, e.g., Sorge & Witteloostuijn, 2004). Based on previous successes or an awareness of the amount of effort involved with design and implementation of a PMS, the temptation is strong to simply start PMS design from scratch (Greenfield), to copy from previous outside assignments or other departments in the organization, or to employ a standardized consulting approach for the design and implementation of performance-measurement systems (Blueprint) (Townley et al., 2003). Such standard consulting approaches tend to focus on strategy clarification and the creation and design of new performance measures, without detailed regard for what is already in place. Existing informal reports typically come into view only after the "ideal" PMS has been designed and set, as part of an assessment of the "gap" between that ideal PMS and already existing performance measures (Medori & Steeple, 2000).

Organizational change is more likely to be successful when it is a process of relatively small change efforts that involve the reconfiguration of existing practices and systems that are successfully in-use elsewhere in the organization, rather the creation of new practices and systems (Abrahamson, 2000). Building organizational capabilities requires adaptation of work processes, reflection upon experiences, and codification of knowledge gained (Zollo & Winter, 2002). In other words, organizational learning is based on *experience accumulation*, and empirical studies have demonstrated the importance of knowledge accumulation for performance (e.g., Reagans, Argote, & Brooks, 2005; West & Iansiti, 2003). Similarly, we propose that building on existing, local experience is an important characteristic of enabling PMS development as well. We expect a development process to successfully stimulate enabling formalization when it fully acknowledges, respects, and utilizes the intellectual capital of lower-level employees' existing practices of and insights in performance measurement.

Experimentation

Experimentation in the context of PMS development involves the first development of a new performance measure and subsequently allowing time to test and refine (in several rounds) its conceptualization, definition, required data, IT tools, and presentation, together with employees (whose performance is going to be measured), to arrive at a measure that is a valid, reliable, and understandable indicator of performance in a specific local context. We propose that a development process that involves much experimentation with new performance measures is more likely to lead to enabling formalization. Fleshing out general goals—the usual suspects of efficiency, productivity, customer satisfaction, etc.—and making them specific and measurable is a "messy" process (Lowe & Jones, 2004). It involves defining measures that reflect strategic goals, that are closely related to the specific operating conditions in a particular setting, that are actually measurable (i.e., the required data are available), and that are presented in a way that employees find understandable. This requires a meticulous, in-depth process of creating a fit between the PMS and the operational idiosyncratic local conditions. The development process requires a close involvement of and cooperation with employees. This is not to say that employees would be the only ones who use the data, but rather that they are the ones who are best placed to judge that their work efforts are validly or invalidly reflected in the performance measures. The making of a performance measure is not likely to be "right" after just one round; it is more likely to be successful if the development engages employees in a process of experimentation, e.g. tinkering with qualitative descriptions, quantitative definitions of measures, the scope of measures, data used, procedures for data gathering, representation in tables and graphs, etc., as well as actual testing to identify unanticipated and often undesirable effects or behaviors that occur in response to the PMS. Even though we emphasize that involving employees through experimentation and building on previous experiences is relevant for improving the *content* of a PMS, this may also contribute to an effective organizational change process (Bourne, Neely, Mills, & Platts, 2003b).

Professionalism

Professionalism denotes an orientation toward learning for the purpose of improving work practices. Such an orientation makes it possible to rely on experience and to conduct experiments within a PMS development process. A higher score on professionalism makes it more likely that employees express satisfaction with earnest improvement efforts carried out within their immediate work environment. Professionalism may be especially stimulated if self-involvement into departmental improvement efforts is made possible. Caldwell, Herold, and Fedor (2004) conclude that employees' motivational orientation, and particularly their "achievement predisposition" (p. 879) predicts satisfaction with perceptions of organizational change. In other words, if an employee is more inclined to improve her work practice, then performance measures are more likely to be seen as positive, stimulating, challenging, and helpful. In sum, we propose that an employee's level of professionalism is associated with a positive attitude towards performance measurement, especially if a carefully evolving developmental approach is taken, aimed at refining and extending a departmental PMS as an instance of enabling formalization.

Research method

This study has been designed as action research. We cooperated with the logistics department of a company in the beverage manufacturing industry, in the period August 2002 through June 2005. We examined in detail the evolution of the department's PMS and the employees' experiences with performance measurement over a relatively long period of time. In this section, we will further introduce the research site, describe how we gathered and analyzed the qualitative data, and outline the survey conducted among a representative panel of the employees of the case department.

Research site

The company has a strong brand name and sells its beverages to both the hospitality industry (such as bars, restaurants, and hotels) and to retail customers that vend to consumers. Customers are both domestic and international. Important conditions for success, according to the company's annual report, are brand strength, product innovation, excellence in production, quality of marketing, balancing stakeholders' interests (shareholders, employees, and environmental concerns), and financial performance. While these factors center on revenue enhancement, cost management is also increasingly important. Competition among supermarket chains has intensified, leading to lower prices for consumers, and increased price pressure on suppliers. The profitability of the company has suffered as a result, and profits, revenues and sales in 2005 were all below their 2004 levels. Furthermore, the company recently made very significant investments in a new manufacturing site, which called for considerable operational cost savings, because it had increased fixed depreciation costs significantly in all departments of the company.

The approximately 150 employees in the logistics department are spread among four sub-departments: purchasing, physical distribution, materials management, and packaging development. The director of logistics and the four heads of the sub-departments form the management team of the logistics department ("logistics management team", LMT). The team also includes the controller for logistics and production, the logistics manager of the hospitality market, and the logistics manager of the international department of the company. The director of logistics reports to the CEO of the company. An organization chart is shown in Fig. 1. The logistics department had been recognized—internally and externally—for its performance, including a prestigious national prize for its customer service and supply chain management.

An overview of some main events investigated during this longitudinal case study is depicted in Table 1. When this study began, the logistics

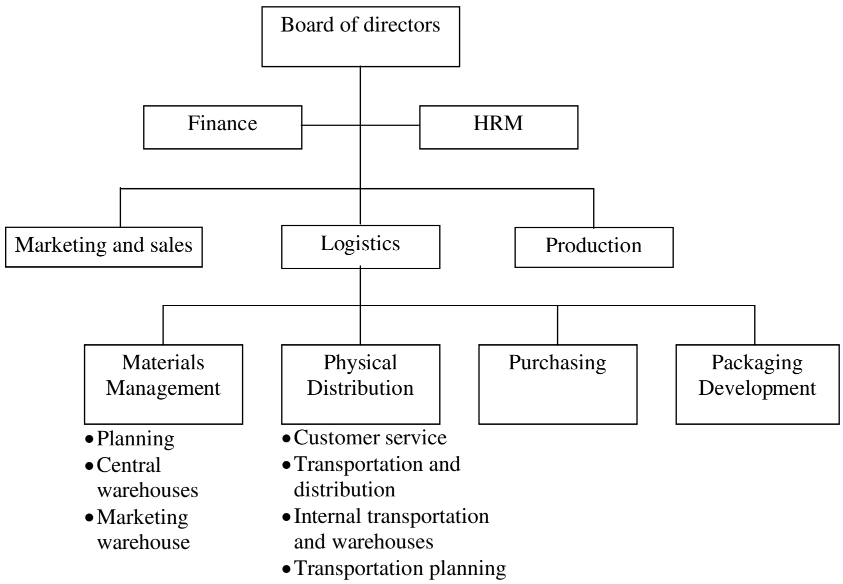


Fig. 1. Organization chart.

department had recently begun to expand their performance-measurement system. They mainly used an indicator called "delivery reliability", but they felt that additional measures were required to provide a more comprehensive picture of the performance of the logistics department in relationship to its objectives. Previously, the mission of the logistics department had been "to coordinate the supply chain in an effective, efficient, and innovative way for providing optimal service to our customers". This had also been reformulated more concretely as four objectives for logistics: number one in customer satisfaction, excellence in supply chain efficiency, continuous

Table 1
Time line of the case study

	Company events	Research
August 2002	Logistics department formulates the need to have more extensive performance measurement	Making contacts with the company and initial discussions about research cooperation
August 2002	Director of logistics voices strong concerns about employee ownership*[a]	
2003		
January	Start developing and implementing new measures with researchers	
August	Positive evaluation of first results and developmental approach	Agreement on longitudinal case study
September	Continuation of design and implementation of performance measures	
December		Start developing survey instrument
2004		
January–May		Developing and reviewing survey instrument
February–March	Logistics department moves to new site	
April	Appointment of new CEO and start of companywide Balanced Scorecard project	
May	Appointment of project leader for Balanced Scorecard project	
June		Pilot of survey
July	Tension from the central Balanced Scorecard initiative*	First survey
August	Start of champions meetings (from all departments)	
October		Discussion of results with LMT[b]
November	Experimenting with a new performance measure for internal transportation and warehouses (continued until April 2005)*	
December	First official scorecards for all departments defined	
2005		
January 2005		Second survey
March 2005		Discussion of results with LMT
March 2005	LMT and middle managers discuss the PMS (new measures and implementation support)*	
May 2005	LMT prioritizes the proposed new measures*	
June 2005	Evaluation of first six months of the official balanced scorecards	

[a] Events market with a * are discussed in some detail in the text with a separate heading labeled "Illustration".
[b] LMT: Logistics Management Team.

supply chain innovations, and to be a professional and learning organization. Explicating these goals stimulated the implementation of performance measures. There was also another reason. In 2002 the company had to impair inventories for about half a million Euros, and therefore it was concluded that inventory risk should be measured regularly. This situation was the basis for the beginning of our cooperation with the logistics department, which provided an opportunity to study in detail the evolvement and actual experiences with performance measurement over a longer period.

In the period between January 2003 and June 2005 the logistics department gradually expanded the PMS to incorporate additional performance measures, to review or delete other measures, and to implement procedures and information systems for producing periodic reports (see Table 2). The development process was strongly influenced by two events: early in 2004 the company moved to a new site and at the same time implemented new information systems that provided new technical opportunities for developing new performance measures. And in April 2004 a new CEO was appointed who initiated a companywide performance-measurement initiative.

We worked especially with three members of the LMT: (1) the director of logistics who reports to the board of directors, (2) the management controller assigned to logistics, and (3) the so-called PMS-champion, i.e., the one sub-department head on the LMT with whom we started this liaison on the basis of her own deeply held professional interest in applying PMS to the entire logistics department within this firm. At the outset we sensed that these leading figures were authentic in their desire to establish a PMS in the form of enabling formalization. They showed keen interest in developing PMS themselves, in cooperation with us as external, university-based experts on both PMS and the human side of organizational change.

Qualitative data gathering and action research

We obtained data through the use of various methods, in the context of action research. Over a period of almost three years we frequently visited the company or met company employees at the university, and qualitative data were obtained through interviews, participation in management meetings, company documents, as well as field notes made by research assistants (see Table 3). While gathering these data, we did not act as neutral observers. The project aimed to assist the company as well as contribute to science. The company participated in this study because they welcomed the unpaid assistance with their development of performance measurement in exchange for offering research access to us. Researchers, research assistants, line employees, supervisors, middle managers, as well as the LMT members were all involved in developing the new performance measures and providing feedback. Over the course of

Table 2
Number of performance measures over time

Period	In use at start of the period	Implemented	Deleted	In use at end of the period	Under construction[a]	Under review[b]
January 2003–June 2003	19	6	7	18	3	1
July 2003–Dec 2003[c]	18			33		
January 2004–July 2004	33	8	6	35	3	5
August 2004–October 2004	35	9	3	41	1	13
Nov 2004–April 2005	41	6	2	45	1	4
April 2005–June 2005	45	2	0	47	6	15

[a] New indicator being developed but not implemented at the end of the period.
[b] Existing indicator being revised during this period.
[c] Data on changes of performance measures during this period were not available.

Table 3
Qualitative data gathered

	Number of meetings	Time (h)
Meetings with[a]		
Employees of Logistics only	8	12
Employees outside Logistics	13	16
Employees from Logistics together with other areas	20	29
	41	57
Number of different employees interacted with		
Logistics	7	
Finance	3	
Production	1	
Marketing and Sales	2	
Other functional areas	7	
	20	

Sample company documents	Number of documents
Research assistants had meetings with 71 different people in 189 meetings, which took over 200 h[b]	
Documents about performance measures in-use in Logistics	13
Documents about performance measures in-use outside Logistics	4
Presentations and notes about developments in performance measurement in the company	8
Minutes of meetings about developments in performance measurement in the company	18
General documents about the Logistics department	11
General company documents	8
Response to panel survey study	4
	66

[a] "Meetings" indicates face-to-face engagements of researchers with members of the case-study organization, either as interviews with one or a few employees, or as active participation in meetings with a larger number of employees. Meetings took place at the research site, with a few exceptions of meetings at the university. Not included are emails and phone calls.

[b] Research assistants also accounted for their meetings. Mentioned are those interactions where they took notes of which the researchers have copies; not included are short informal discussions (the research assistants worked on-site), emails, and phone calls.

this study, seven master students in industrial engineering or business administration worked full time for a period of six to eight months as an intern at the company, in partial fulfillment of their MSc. They produced monthly reports of actual outcomes of performance measures, they carried out the survey, and they worked with employees in developing, evaluating, or refining various measures.

The qualitative data have been analyzed through a process of reflection, and going back-and-forth between the data, the literature, and the company. The research file was organized on the basis of a table that listed the interactions with the company and that contained about 275 rows. On each row the following data were recorded (when applicable): date, who participated, sub-departments involved, topic, duration, reference to meeting notes, description of (and reference to) company documents received, reference to researchers' input for the meeting, code for meeting in person, code for researcher or assistant, code for diversity of meeting participants. This table allowed easy retrieval of specific data when certain questions or ideas emerged in discussing or writing about the research. It was also the basis for summarizing the time line in Table 1 and the data gathering in Table 3. The data were used to write summaries to pull together different events and different kinds of data, and to start reflecting on what happened, and to focus on events that seemed most interesting. Parallel to gathering the data and writing the summaries, we reviewed more literature, discussed the study with other researchers (informally, as well as through presentations in workshops and conferences), and wrote (and rewrote) the paper. This connection with theory guided not only the analysis of the data, but also the gathering of data, and it led to follow-up discussion or clarification with the company. Also, a draft version of the paper was discussed with managers of the case study organization. And vice versa, interaction with the company guided the search for new literature, or informed discussions with academics.

A potential issue of action research is that the researcher may selectively look for empirical evidence and guide the research process with a bias towards the expected findings (Atkinson & Shaffir, 1998). However, there are several countervailing effects that limit such a bias, which were also prominent in this study (Labro & Tuomela, 2003): The length of the research process and access to all kinds of data provides many different "pieces of the puzzle" that provided different kinds of empirical evidence which need to be understood as a whole. Furthermore, members of the case study organization expect results that are of practical relevance and this provides an incentive for them to be involved and to spend time with the researchers. Also because of the potential impact on their work, organizational members are engaged, challenge ideas, and provide feedback on results. Furthermore, this type of action research allows an empirical test of ideas implemented in an actual organization. Organizational members are likely to be cautious about trying interventions they deem unsuccessful or otherwise undesirable. Researchers cannot easily persuade people to implement what they consider to be a bad idea; and when an idea that seemed good actually works out poorly, that will become obvious in the empirical data. In short, the objective of making actual changes in real organizations counters researchers' biases, because of the active involvement of organizational members and the empirical facts resulting from implementation.

Panel survey

A survey was conducted twice, once in July 2004 and once in February 2005. On both occasions we approached the same respondents, and we assessed also the same variables, with the same or a slightly improved version of the questionnaire. Hence this part of the study is labeled the "panel survey". The timing of the panel survey within Logistics coincided with the start of the companywide initiative to implement a balanced scorecard, and the two surveys provided information on the initial attitude towards performance measures and the situation about six months into the initiative.

We first asked the four sub-department heads within logistics to come up with a list of potential respondents who would be representative for their sub-department in terms of their attitudes towards PMS. The LMT reviewed the lists and made a few changes of prospective respondents, which were subsequently approved by the nominating sub-department heads. Members of the panel had to have been employed in their sub-department for at least one year and not temporary employed. In addition, all sub-department heads were included since they were crucial in the PMS process. Moreover, the number of panel respondents per sub-department had to be proportional to the size of each of the four sub-departments.

In the first data wave, we received the completed questionnaires from all of the 42 selected respondents. In the second data wave, we got the data from 39 of the same 42 respondents plus one new participant. This attrition was due to illnesses, and one employee on the panel had left the company.

For the process of data collection, research assistants requested the participation of panel members with a letter (signed by the director of logistics) and verbally during several team meetings in which they explained the purposes of the survey and the role of the panel and allayed participants' concerns about the confidentiality of the data. To ensure a high response rate, the research assistants made appointments with all members of the panel to have them fill out the questionnaire during an on-site interview. The assistants also wrote down other PMS-related comments respondents made during the meetings. In the second survey appointments were made only with those respondents on the panel who were expected to be uncomfortable with completing the questionnaires by themselves.

Confidentiality was a key consideration during the panel study. Given that some sub-departments were rather small and that panel membership was known by the LMT, we promised the respondents explicitly and repeatedly that results would never be reported at the sub-departmental level, but only at the aggregate level (i.e., for the whole logistics department). Furthermore, the completed questionnaires were filed outside the company (at the university) and no one at the company possessed the list that linked respondent numbers to respondent names. Procedures to guarantee confidentially of the data were emphasized in all communications with research participants, and thus participants appeared comfortable enough to provide frank answers and comments.

The measurement of the panel survey data will be discussed in the remaining part of this section.[1] The measurement scale for the dependent variable *Attitude toward performance measures* was developed expressly for this study. Its items are described in Appendix. The variable reflects the perceived usefulness of performance measures that are reported concerning the respondent's subdepartment within the logistics department. In the second administration we added the variable *Ambition level in two years*. Using the same items, participants were asked what the situation with respect to performance measurement should be two years into the future.

The measurement of *Professionalism* was also developed expressly for this study. Its items are also described in Appendix. We refer to this new construct informally as being improvement-oriented on the job. Formally, professionalism refers to the degree to which individual employees behave in a way that shows commitment to both their profession and their current organization, through efforts aiming explicitly to upgrade or improve the quality of the work carried out. Sample items are: "I learn every day at work"; "I always contribute to new ideas at work". The answering possibilities of these Likert items range from 1 (very much disagree) to 7 (very much agree). The origin of the professional attitude questionnaire lies in efforts carried out by Swailes (2003) who in turn relied explicitly on measurement efforts of Hall (1968) and Snizek (1972). In our study we defined the questionnaire items entirely on an individual employee level. Deviating from these previous professionalism scaling efforts, we made all items refer to solely one's own current job and not also to one's profession, other professions or professional colleagues. Because of a lack of validity of broad measurement scales, Swailes (2003) called for "reconceptualising professionalism in terms of process rather than structure" (p. 103), to which we contributed in this study through the formulation of the survey questions.

In the survey, a number of variables regarding the task environment were also included. This made it possible to investigate the association between professionalism and the attitude towards performance measures while controlling for these other variables that could also affect the attitude towards performance measures. *Leadership style* was measured using a subset of 10 MLQ 8Y items of transformational leadership style (Bass & Avolio, 2000). MLQ refers to the currently most used valid questionnaire for assessing leadership. For *Team trust* the scale was comprised of the seven items employed in a German study by Baer and Frese (2003), and based on the work of Edmondson (1999). The scale for measuring *Work pressure* was comprised of 14 items and is taken from Stanton, Balzer, Smith, Parra, and Ironson (2001). *Work satisfaction* was measured in a way in which respondents had to write down three numbers totaling 100%. They were asked to note the percentage of time they felt, on average, "satisfied", "unsatisfied", and "neutral" about their current job.

Results: developing an enabling PMS

In this section, we will present the empirical material to explore the developmental approach that fostered the enabling nature of the PMS in the case study company. First, we will present results that suggest that employee attitude toward performance measures in the logistics department was quite positive, and this is based both on the survey and the qualitative data. Then we will explain this positive attitude through the propositions outlined above in propositions about a developmental approach for enabling PMS section.

[1] Please contact the first author for more details about the research instrument.

A positive attitude toward performance measures in the logistics department

Employee attitudes toward the performance measures used in their sub-departments were quite positive; the means from two waves of questionnaire deployment are in Table 4. Note that the reliabilities of the survey questionnaires ranges from satisfactory to good, as measured by Cronbach's Alphas. The average scores on the variable *Attitude toward performance measures* were 5.2 and 5.4 (for the first and second data wave, on a seven-point scale). In the second administration we also asked participants what the situation with respect to performance measurement should be two years later, using the same seven items and the same seven-point answering scale (*Ambition level in two years*). The average score shown in Table 4 is 6.2, which is higher than their assessment of the current situation. This suggests that on average employees were ambitious in terms of performance measurement, and this may also be taken as an indication of a positive attitude toward performance measurement.

The qualitative data provide further evidence of positive attitudes toward performance measurement. Particularly, significant was a meeting with the logistics management team (LMT) and middle managers in logistics (i.e., managers who reported to the members of the LMT, planners, and shift leaders) in March 2005 (to be described in detail below). We will describe that there were tensions: the middle managers wanted more performance measures to support them in their work. To the extent that they could not develop and implement these themselves, they needed resources outside their teams (such as time from experts in the controller's office), and the prioritization of such resources was debated. We take this as another indication of positive attitudes toward performance measures, as the PMS within logistics was clearly being perceived as enabling formalization.

Illustration: LMT and middle managers discussed the current status and priorities for further development of the PMS

The PMS-champion and the controller presented the history and current state of performance

Table 4
Survey constructs (reliabilities, descriptives, correlations)

First data wave, $N = 42$	Cronbach's Alpha	Mean	SD	1	2	3	4	5
1. Attitude toward performance measures	.816	5.169	.818	1				
2. Professionalism	.813	5.304	.606	.433**	1			
3. Leadership style	.862	4.848	.891	.137	.275	1		
4. Team trust	.729	5.762	1.153	.520**	.208	.302	1	
5. Work pressure	.803	3.628	.741	−.224	.110	.041	−.409**	1
6. Work satisfaction	.868	6.000	1.653	.316*	.238	.175	.543**	−.421**

Second data wave, $N = 40$	Cronbach's Alpha	Mean	SD	1	2	3	4	5	6
1. Attitude toward performance measures	.906	5.430	.797	1					
2. Ambition level in two years	.881	6.186	.627	.681**	1				
3. Professionalism	.878	5.317	.587	.385*	.439**	1			
4. Leadership style	.914	4.463	1.050	.219	.311	.365*	1		
5. Team trust	.655	5.500	.973	.205	.194	.351*	.326*	1	
6. Work pressure	.870	3.586	.852	−.281	−.105	.019	.072	−.363*	1
7. Work satisfaction	.909	5.783	1.924	.177	.124	.242	−.026	.490**	−.485**

For completeness we also conducted a principal components analysis. All measurement items loaded on their expected factors for the independent variables Leadership style, and Team trust, but not always for Professionalism, Work pressure, and Work satisfaction. However, because of the very small number of observations, we maintained all items for further analyses.

* Correlation is significant at the 0.05 level (two-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

measures in a meeting of the LMT together with middle managers in logistics, in March 2005 (see Table 1). They contrasted performance measures that provide insights into whether the logistics department achieves its medium and long-term objectives versus performance measures that an employee in logistics may need to be successful in his or her daily work. The PMS-champion and controller described their ''dream'' situation: ''middle managers can and want to develop performance measures and produce the reports on these themselves in order to better manage their processes''. The director of logistics explicated this further as ''inventing it yourself; getting the data out of systems or recording these oneself. Is this a dream or a nightmare? Is this something we share? Or do you just always want to call the controller, who should realize this for you?'' This remark stimulated a lot of discussion. It seemed that the idea of developing their own measures and reporting on these was supported. ''There are enough opportunities with the new systems, and sometimes you discover only after a while what you can get out of these'' one of the participants commented. Middle managers further remarked that they felt they needed to be fully involved in the development process of new measures, and it should be made easy for them to actually generate the reports.

The discussion centered on the way in which resources for PMS development were allocated between performance measures for the LMT and those for use by middle managers in logistics. At first, practical issues with performance measures were mentioned, such as that currently, for many performance measures, the reporting involved manual activities that required too much time and that could cause errors. The middle managers said better information systems and tools were needed. They also expressed a desire for specific new performance measures to be implemented. During this discussion, it was brought forward by the middle managers that support from people in the controller's office and from the research assistants was required. It became clear that the resources for developing and implementing all the desired new performance measures would have been severely overstretched, and thus not all wishes for new performance measures could be supported. The controller stated that he wanted to allocate the resources to performance measures that involved significant financial risk, which was difficult to control without such measures. Hence, employees in logistics had to develop and implement some measures by themselves. While there seemed to be much support for the idea of developing their own measures, it was also discussed that in some situations this was considered too difficult, and specialized involvement from the controller's office was needed. The middle managers argued that their employee voices should be heard and their requirements should be supported.

The PMS-champion mentioned a particular measure and said that she had a real dilemma about it: ''I agree with [the controller] that we should focus on the strategic measures for logistics, but I also feel that this is a really important measure within our department''.[2] One of the managers said: ''If I as a middle manager ask for a particular performance measure, I think you should say 'yes' right away, because then I really need it, and otherwise you will not get any support [from us middle managers for performance measurement initiatives]''. The controller responded by saying ''that is simply not always feasible, our time is limited''. To this the manager responded provocatively: ''so if I am held accountable for something, I get no support, but if the logistics management team is held accountable for something, then there is support?'' Clearly, people held different opinions about how dependent they were on specialized support. Another middle manager commented: ''But these practical issues have never stopped us from going forward with implementing performance measures. .... And if you look at our performance measures in [the sub-department], we designed and implemented these almost completely by ourselves''.

Later in the meeting, four groups discussed ideas for new performance measures and presented these.

[2] The word ''department'' was used in the company also to refer to the four sub-departments within logistics. ''Sub-departments'' is a term we use in the paper for clarity, but we write ''department'' in quotations and in the questionnaire in Appendix, as this term was actually used in the company.

As a follow-up, the sub-departments within logistics were asked to think about these measures further and come up with proposals. After this meeting, the sub-departments proposed 16 new performance measures in total. This list was input to a prioritization meeting of the logistics management team in May 2005. The team prioritized the new measures. Moreover, it was concluded that in the future middle managers should be involved more in further developing the PMS. One of the members of the LMT reflected "we have been very much top down, authoritarian with our kpi process".

Just to illustrate the dilemma for prioritization further, it is helpful to look in more detail at one of the performance measures that the PMS-champion referred to above. It shows that middle managers within logistics had specific ideas about new performance measures. Toward the end of 2004, one of the warehouses was nearing capacity, running the risk of an overflow situation. A group of people in logistics had developed solutions for this problem and subsequently wanted to monitor the effect, to see how the utilization of the warehouse was developing over time. This was a simple graph showing on a daily basis how many pallets were stored. This total should be around 600 pallets maximum, and serious storage problems would result if it rose above 800. This inventory monitoring report had been made available, with the help of both the controller's office and the research assistants. It may appear very simple and easy to set up, but it took a couple of days, preparing the SAP downloads and setting up the Excel sheet to develop and implement this report. A similar type of performance measure was now (in March 2005) needed in another warehouse, where all carton products (labels, boxes, etc.) were stored. Note that outside storage was not an option for these items. Again, a graph would be needed showing the warehouse utilization on a daily basis, but the new graph was a bit more complex. It needed to indicate warehouse utilization disaggregated into different types of storage. Setting this up would take several days of effort by the controller's department, who claimed that time was not available for designing and implementing this performance measure. This is an example of a new performance measure that middle managers wanted and that stirred debate on the prioritization and resource allocation for the development of performance measures.

These examples of refinements to the PMS that were initiated by employees to better support their work practices suggest that the nature of the PMS in logistics was predominantly enabling rather than coercive. Can the enabling nature of the PMS be understood based on the development process that had shaped this PMS? In the following sections, we will report on three characteristics of the process: (a) professionalism, (b) experience-based PMS development, and (c) experimentation with new performance measures. We will also explore how internal transparency (d) was important for encouraging enabling formalization.

Professionalism of employees established the basis for PMS development

This beverage manufacturing company was recognized for its professionalism by winning a prestigious national prize in the retail beverages category. In this annual contest, supermarket chains assessed their 90 largest suppliers in terms of three criteria: account management (the quality of the sales team), trade marketing (the quality of the sales support for the supermarkets), and supply chain management (the quality of the logistical processes). For these criteria, the company had won the highest score of all beverage suppliers assessed (water, soft drinks, beer, wine and spirits) for three years in succession (2002, 2003, and 2004). This suggests that the logistics department operated at a high professional level.

We proposed that professionalism contributed to a positive attitude towards PMS. Regression results are presented in Table 5. The dependent variables are *Attitude toward performance measures* (both data waves) and *Ambition level in two years* (second data wave) with respect to performance measurement. As shown in the table, the coefficients for the variable *Professionalism* are statistically significant, and they are considerable (.521, .518, and .420) and larger than the coefficients for all other independent variables.

These results show that a high level of professionalism is a key characteristic of a development

Table 5
Regression results

Dependent variable	Attitude toward performance measures (first data wave)	Attitude toward performance measures (second data wave)	Ambition level in two years (second data wave)
Intercept	1.498 (1.206)	3.843 (1.322)*	4.061 (1.049)*
Professionalism	.521 (.191)*	.518 (.231)**	.420 (.184)**
Leadership style	−.086 (.129)	.097 (.129)	.118 (.103)
Team trust	.321 (.118)*	−.062 (.158)	−.038 (.125)
Work pressure	−.108 (.170)	−.324 (.166)[a]	−.113 (.131)
Work satisfaction	−.023 (.081)	−.018 (.079)	−.004 (.063)
# observations	42	40	40
R^2	.397	.250	.238

Unstandardized coefficients and (Standard errors) tabulated.
We also estimated eight alternative specifications of these regression models, always including Professionalism plus various combinations of a number of the other independent variables. The coefficient for Professionalism was positive and significant at least at the .05 level in eight cases (first model), in seven cases (second model), and in eight cases (third model) (results not tabulated).
[a] Coefficient significant at .059.
* Coefficient is significant at the .05 level (2-tailed).
** Coefficient is significant at the .01 level (2-tailed).

process with high employee involvement. It enabled employees, together with professionals from the controller's office, to experiment with performance measures and to gradually expand and refine the PMS on the basis of learning from experiences. We will elaborate now on the other two characteristics of the PMS development process.

Experience-based PMS development process built on existing measurement practices

Experience-based PMS development refers to identifying and building on local experiences with performance measures during further rounds of refinement of the PMS. Note that ''experience-based'' points to the capturing of experience for guiding development at the level of the performance measurement *system*; in the following section ''experimentation'' will be discussed at the level of *single measures*.

Our qualitative results indicate that the logistics department has been following an experience-based development process. Table 2 shows the development of performance measures in logistics between January 2003 and June 2005. The total number of performance measures increased from 19 to 47 measures. New measures were being added constantly, while other measures were removed. Still other measures were reviewed and updated and re-implemented. This situation is not reflective of a PMS initiative that is first designed, separately implemented, and then reviewed, for example annually. Rather, the image is that of a more ''organic'' PMS that is constantly growing, being reviewed, and being pruned—a continuous tinkering to make it better. It is consistent with processes of incremental improvement, based on experience gained (Abrahamson, 2000; Zollo & Winter, 2002).

We will illustrate below that especially the LMT attached importance to an experience-based process for bringing their PMS further. In sum: the company's top-management wanted a common format and approach to performance measurement in all departments of the organization. They labeled this ''the balanced scorecard''. Top-management initiated such a performance measurement project during the period that we studied. Tensions between the central initiative and the local experiences in logistics could be observed. While logistics was already the most active department in terms of performance measures, the central initiative was at several times perceived by the LMT members as something that could disrupt rather than foster their ongoing PMS activities.

The central initiative placed performance measures high on the agenda, and this priority status could have possibly provided momentum to help the LMT to move their PMS initiative forward. However, by July 2004 the LMT members were concerned about what the central initiative would mean for the PMS they had so carefully developed with their employees over the last years. We understand this as another indication of the importance of an experience-based development process. There was tension, because the people in logistics worried that this top-down initiative would not reflect their experiences and would not allow time to experiment with and adjust performance measures.

Illustration: tension from the central balanced scorecard initiative vis-à-vis the PMS in logistics

The new CEO was appointed in April 2004, but he had already been a member of the board for several months to get to know the company. He conveyed his emphasis on performance measures right from the start: he attended a meeting with the LMT in October 2003, during which performance measurement was a main point on the agenda. He made it clear that he considered performance measurement to be very important and he wanted more of it, throughout the company. He wanted a system to be implemented quickly: defining the measures, setting (and "freezing") the targets and tolerances. In that meeting, he also said he wanted to show performance as a traffic light that would show red when measures slipped below their target, and in which case the manager responsible for a particular performance measure would need to prepare an action plan for presentation to the management team. He spoke of this in terms of management by exception, whereby the performance report should be used to select issues that needed to be discussed. It was clear that once he assumed the chairmanship, performance measurement and reporting were going to be matter of high priority and a focal point of top-management attention.

In June 2004, the new CEO announced the new balanced scorecard project in the company. A balanced scorecard at the board level—so for the firm as a whole—was being formulated and each of the different departments in the firm, of which logistics was one, had to devise a business balanced scorecard for their area. A companywide project leader—an experienced internal manager at the director level, reporting directly to the CEO—was appointed. This initiative soon created anxiety within the logistics department. The manager in logistics who had become the main proponent of performance measurement (the PMS-champion, but that title was not yet used at that point in time) called a meeting with the researchers, also on behalf of the controller. She informed the researchers about the central initiative and explained that the project leader was talking with other companies and consulting firms. She expressed concern that new performance measures would be determined in a top-down fashion by the central initiative, led by consultants. The concern of the logistics group was that their long-standing and ongoing PMS work would now be disturbed by the top-down mandated initiative. They feared that their system would have to be changed to comply with the new, companywide balanced scorecard framework. They were not against more performance measurement—on the contrary, they had already implemented performance-measurement initiatives in logistics—but they feared that a consultants-led project would be started with top-down proposals for new performance measures, and they expected that this would leave less room for what they had developed so far, which had a close fit to local work practices.

During the second half of 2004, the project leader of the central balanced scorecard project visited several other companies to discuss their experiences with the implementation of a PMS. He concluded that although these experiences were very diverse, it was clear that effective PMS development take several years, and are more successful if developed bottom-up and from within the organization, and that an organization should simply get started and develop things further as it goes. He also had gained the impression that when performance measures were part of the incentive system, there was a significant risk of manipulation. He wanted to use these insights in the scorecard project for which he was now responsible. However, there were tensions and he said in a consultation with us, that in the eyes of the new CEO "things are going far too slowly". The project leader established a group of "champions" in August 2004. There was one person from each department within the company who was the most enthusiastic proponent of performance measurement and who was leading departmental initiatives to develop it further. With one exception, these were not controllers, but functional managers.

The deadline for the first version of all departmental balanced scorecards was November 1, 2004, and it was postponed until December 1, 2004. However, not all departments met the second deadline, whereupon the CEO put a traffic light in the central entrance hall of the company's premises, with the signal showing red. It was there for a couple of days. There was no sign or other explanation of why it was there, but managers soon found out it was there to signal that scorecards really would have needed to be completed. The intervention provoked quite some discussion, some of which we witnessed in a meeting with the LMT. Members of the team acknowledged that it was unfortunate that the deadline was not met, but they did not like the traffic light intervention. They complained that the efforts that other departments were undertaking were not facilitated, in the sense that the practical support for implementation was lacking. The logistics' PMS-champions objected: "We have said to the project leader on several occasions that this can only be done if it is facilitated in a practical way, but nothing has happened". The logistics director commented that such comment did not reach the company's directors meeting when the balanced scorecard initiative was discussed.

The balanced scorecards were implemented in the middle of December 2004, and they were evaluated internally and changes were proposed on the basis of the first six months of experience. The balanced scorecard initiative was perceived positively, according to the company project leader. In a recent strategic planning meeting with the top-35 managers of the company the balanced scorecard was often mentioned as a positive development. However, the companywide project leader also felt the process at the time was fragile: "If I would stop [leading this project] now, then it would collapse. So, apparently it's not yet deeply ingrained". Balancing the top-down pressure from the chairman "who simply wanted to have this" and letting the balanced scorecard being developed bottom-up was very important (and sometimes very difficult), according to the project leader.

Yet, an event in the middle of 2005 again pointed to the importance of allowing an experience-based process. A consulting firm that the company had engaged on another project had devised another, different balanced scorecard proposal for the entire firm. It did not build on what had been developed thus far (i.e., Greenfield approach) and it was based on what the consulting firm considered to be best practice in other companies (i.e., a Blueprint). The balanced scorecard project leader considered it a very serious mistake if the firm were to adopt the blueprint proposed by the consulting firm and to present it at the next top-35 meeting: "What are people such as [the logistics PMS-champion] supposed to think if this hangs on the wall in our next top-35 meeting?" He emphasized that the proposal did not do justice at all to what the company had developed by now, "which is so specifically modeled to our situation".

Experimentation with new measures

Experimentation refers to the process of tinkering with a single performance measure while designing and implementing it. This means that design and implementation are interconnected, because the design (from conceptualization to fine-tuning the presentation) is partly done with real data, after measurement and reporting on the new measure have already begun. New performance measures are hardly ever "right" straight away, and by allowing adjustments the reliability and validity of the measure can be improved if taking into close consideration the context where the new performance measure is actually in use. In other words: both conceptual and detailed implementation issues of performance measures are crucial for their effectiveness. Employees typically possess key, yet tacit, knowledge that is required to further refine performance measures. In performance-measurement development "the devil is in the detail", as illustrated below, with respect to a new efficiency performance measure.

Illustration: experimenting with a new performance measure for internal transportation and warehouses

We will illustrate experimentation with a new performance measure for the sub-department of "internal transportation and warehouses". The activities took place from November 2004 until April 2005. The main activity of this sub-department was to store finished goods in the finished goods warehouse (it was brought from production to the warehouse by an automatic transportation system) and to load and unload delivery trucks, using forklift trucks. The workload for this activity was unevenly spread throughout the day. For managing efficiency, planning the number of operators per shift was important, and also making sure that the operators were carrying out the work quickly and that certain preparatory activities were done during idle time. A new performance measure was thus needed for efficiency purposes. (Observations of forklift-truck drivers gave a rough indication of efficiency, but the managers and first-line supervisors wanted more factual data to complement these.) Based on existing ideas within the sub-department and the controllers' office, and on discussions with employees in the transportation sub-department, the performance measure was defined as the number of "transports" carried out per labor hour. For example, one transport could be to pick up a pallet from the automatic conveyer belt and to bring it to a particular location in the warehouse.

Measuring the number of transports was feasible, because each transport was issued by the warehouse management system to terminals on the forklift trucks. Measuring the number of labor hours spent was also possible using the warehouse management system, and so the ratio of the two was readily available. However, not all activities that needed to be carried out were issued by the warehouse management system. For example: particular types of pallets needed to be rotated 90° before they could be transported; the forklift-truck drivers sometimes needed to move lorries for loading and unloading; containers for international destinations needed to be closed and sealed. An initial list of more than 30 of such side activities was prepared and a copy was given to each forklift-truck driver to estimate the time spent on these activities, and to add new activities to the list. The final list of side activities was compiled and the workload for these activities was also expressed as a number of transports (the numerator of the performance measure). A manager in the sub-department "internal transportation and warehouses" calculated the performance measure, by downloading data from the warehouse management system and preparing the reports using Excel.

The target for the new performance measure was an engineered target, as this was based on the detailed design of the processes in the warehouse. However, not all assumptions underlying this design were initially met, as some processes were carried out differently, and this led to an adjustment of the target. The presentation format was designed such that it included, in weekly numbers, the absolute number of transports and the number of transports per labor hour (both per week and cumulative).

After the initial measure and report had been implemented, weekly evaluations were conducted with the sub-department's manager and the team leaders. They provided relevant and detailed feedback, in particular regarding the way in which side activities were included in the calculation of the performance measure. For example, for some side activities it was decided not to estimate the amount of work involved every week, but rather include a more general estimation. It was also discussed whether the performance as measured made sense and appeared valid from the team leaders' perspective, which was found to be the case. The weekly frequency of reporting the measure was found useful. Also the presentation format was discussed, because it was quite a complex chart. Adjustment to the chart were made, but also the term "overcapacity" that was initially used was replaced by "theoretical utilization" and the chart's original label "efficiency [sub-department]" was changed to "transportations per labor hour in [warehouse name]".

Similar experimentation activities were conducted for other performance measures. The effect of experimentation was not only that behavioral effects such as commitment were improved. More to our point, experimentation played a vital role in arriving at a performance measure that was more reliable, valid, and understandable in the context. The people who were responsible for the PMS (from the controllers' office and research assistants) obtained an in-depth understanding of the operational processes the PMS were supposed to capture. These specialists needed to obtain an intimate familiarity with the operational processes, and the operational managers needed to understand the details involved in actually translating these processes into quantitative performance numbers.

In sum, in the above and previous sections we reported findings showing that a PMS is more likely to be seen as a constructive, enabling type of formalization, rather than a negative, coercive form of control, if it is developed incrementally such that the members of the organization can gain actual experience with using performance measures, reflect on this, and draw conclusions to develop the system further. We observed ongoing activities such as reviewing and revising existing performance measures, brainstorming about possible new measures, experimentation with new measures, adding some new measures to the PMS, and dropping some existing measures. There was not a specific point in time when the performance-measurement system was "ready". It was also found that the throughput time for actually implementing a new measure was considerable and could easily take half a year to one year. Furthermore, it was crucial to look in detail at existing measures at the start of each measure development. New measures could only be developed after understanding and using as much as possible from what was already in place such as the precise definitions of existing measures; the various rationales behind these; the data used; the limitations that people experienced with the existing measures; the ideas that people were working on to improve the existing measures; and information system changes that could impact existing reports. Hence, neither a Greenfield nor a Blueprint approach was taken. The developmental approach stimulated the inter-functional exchange of knowledge, which was required for a reliable, valid and understandable PMS. It created also transparency of the system from the employees' perspective (and transparency of operational processes for the PMS specialists). Therefore, we will explore PMS transparency in the next section.

Internal transparency was emphasized throughout the development process

Transparency, in combination with flexibility, in the context of performance measures, means that employees (whose performance is going to be measured) are highly involved in operating and managing the PMS as organizational technology. Transparency and flexibility imply that the performance measures are understandable to employees, something they have hands-on-experience with, and something they can influence to make it workable for them. In the studied company, performance measures were not owned by, nor understood solely by the technical specialists in the finance and accounting function. Instead, employees had been an integral part of the development of the measures from the outset. They were in some cases even managing the system after it had been implemented. The operational managers themselves were trained with the information system tools to record data, to pull together data, to create performance reports, to review and revise definitions of performance measures, to change graphical representations of performance reports, etc. The director of logistics was a strong proponent of non-accounting ownership of the PMS, and we will elaborate on this below.

Illustration: director of logistics voices strong concerns about employee ownership of the PMS throughout the study

The director of logistics was very outspoken on matters of what could be considered internal transparency. Already in the first meeting during this study, in August 2002 (see Table 1), he emphasized that he wanted the employees rather than the controllers to be responsible for reporting performance. "If people are not going to take the effort to do the measurements and make the reports, it probably means it's not essential to do them". Whether new measures would actually be implemented was a kind of relevance test in his mind. In his view, when confronted with performance measurement, employees would ask themselves "Do I feel responsible for this?", "Does it help me?", and that the answers to these questions would determine their attitudes and level of cooperation.

The director of logistics expressed these concerns at the beginning of the study, when he stated that the emphasis should be on creating new performance measures for use within the four logistics sub-departments, and that less emphasis should be given to measures for logistics as a whole (for use in the LMT) or for reporting the performance of logistics to the board. And towards the end of the project (in May 2005), when discussing the priority for further development (after the meeting with middle managers in logistics, March 2005, described above), he made forceful comments reflecting concerns for empowering middle managers. As described above, in March 2005 the four sub-departments within logistics had proposed a list of 16 performance measures they wanted to implement, and there was a meeting of the LMT about prioritizing those measures and discussing more generally how logistics wanted to move forward with their performance-measurement system. In that meeting, the question was raised whether, if needed, the LMT was willing to allocate some of the resources to support performance measures for lower-level managers within logistics, and if they wanted to invest in enabling these managers to implement and generate performance measures themselves. Such investment would include, for example, buying and implementing additional IT tools, providing training, allocating time of people in the controller's office. Some members of the LMT supported this, but also had some reservations: e.g., "Only if I also think it is relevant for our department". The logistic director intervened: "You cannot give a conditional 'yes', no 'yes, but'. You cannot say 'no' to this". He stated that, as a principle, he wanted to support the information requests from lower-level managers, and subsequently there could be a need to prioritize. "And those that are selected, we will certainly enable". "If we then say 'these [particular performance measures] are the most important ones and these we will facilitate', then managers should define what it entails and come up with a project plan for each [performance measure]".

In sum, during various discussions (the meeting in August 2002 with the middle managers, and later in the LMT) it became clear that different PMS requirements existed for the LMT member and for middle managers in the logistics department. There were no inherent conflicts between these different requirements, because the PMS was conceived of and implemented as something that supported different managerial levels—the enabling intent and nature of the PMS was unmistakable. However, in practical terms, there was a conflict in the sense that resources for PMS development were limited, and choices had to be made regarding whose PMS requirements were going to be implemented first. This observation illustrates a key advantage of transparency due to heavy involvement of employees throughout the development process: as employees were more involved and better enabled (such as provided with IT systems for PMS development), internal transparency increased and dependency on specialized resources for further PMS development was reduced.

Discussion

This study of performance-measurement systems (PMS) in operations focused on identifying a development process that is likely to lead to a PMS that employees regard as useful for them; something they want to help develop, and not exclusively as a control device for senior management. Which characteristics of the development process contribute to an enabling PMS? How can a PMS be developed as enabling formalization and not as coercive formalization? Our research was conducted as a longitudinal case study of the logistics department in a medium-sized beverage manufacturing company, from August 2002 through June 2005. Qualitative data were gathered, as well as two waves of survey data.

We found that *Professionalism* was significantly related to positive attitudes toward performance measures, based on the survey data. The qualitative findings point to professionalism as a force that can be mobilized through a development process that is experience-based and allows for experimentation. *Experience-based* characterizes a development process that builds on existing skills, practices, and know-how of involved employees, in order to enrich the PMS step-by-step over time. Key qualitative findings supporting the importance of this experienced-based characteristic were revealed when a centrally initiated balanced scorecard initiative threatened to overrule the experienced-based approach that had been followed thus far within logistics. *Experimentation* with PMS improvements concerned deliberate employee efforts to test, review, and refine conceptualizations, definitions, data, and presentations of new performance measures. Key findings supporting the importance of experimentation pertained to the way in which specific new measures (such as for warehousing and internal transportation) were developed. Furthermore, we found that *transparency* contributed to an enabling PMS. This became apparent from the internal discussions on ownership of the performance measures. Local transparency in the context of performance measures was stimulated by deeply involving operational managers in the conceptual and practical development of such measures, but also by making them, rather than people in the controller's office, responsible for periodically "calculating" and reporting these measures.[3]

The theoretical framework developed by Adler and Borys (1996) was key to our study of effective organizational change towards an enabling PMS. We developed their framework further, based on other literatures and the empirical data, in the specific context of PMS for operations. We proposed three characteristics of a development process that is likely to result in an enabling PMS, and we show a departmental episode where enabled formalization took place in form of a much-expanded PMS. Its development process was characterized by experiential inputs, experimentation, and a high degree of professionalism on the part of individual employees. Furthermore, we found that the "norm" of transparency repeatedly and consistently voiced by the departmental director was a force that contributed to the enabling PMS we witnessed in this setting of a logistics department.

Conclusions

This study provides several main contributions. First, it increases our understanding and appreciation for a developmental approach leading to an enabling PMS. We demonstrate that building on *existing performance-measurement experience* of employees, as well as their professionalism, and allowing *experimentation with measures* contribute to the enabling nature of the PMS. Design and implementation appear interrelated, because design is partly conducted while obtaining an empirical understanding of how performance measures are being used within their actual operational context. An experienced-based process and experimentation are not used to manage resistance to organizational change, such as to create commitment, or to make people feel that they are taken seriously (Piderit, 2000). Rather, an experienced-based process and experimentation serve to involve employees in such a way that their knowledge is mobilized to design a more valid, reliable, and understandable PMS in their specific local context. We pointed to the importance of *professionalism* of employees as a condition for a development process characterized by being experience-based and containing experimentation.

Second, this study provides a possible explanation for *why* a developmental PMS approach as described in our case study may establish enabling formalization. Given the fact that a developmental approach to PMS evolution engages all personnel whose performance is being measured, it may compensate for the inherent *incompleteness* of performance measures. This study connects accounting considerations about the completeness of performance measures in operations (Chapman, 1997; Lillis, 2002) with the ideas proposed by Adler and Borys (1996) regarding design principles for

[3] After this study, around January 2006, it was formally decided that the reporting of most performance measures would be the responsibility of particular managers in logistics. The controller's office would conduct audits on these measures, and with the IT department they had to provide skills and tools to managers in logistics. The controller's office was responsible for reporting on the measures that were included on the scorecards for the LMT and the company's top-management.

enabling formalization. A PMS reflects performance on a variety of dimensions, such as efficiency, productivity, quality, and responsiveness. However, it remains difficult to develop a technically complete PMS that fully reflects the dimensions of operational performance, that contains valid measures on all these dimensions, and that includes targets that reliably capture the tradeoffs between opposing performance measures (Chapman, 1997; Lillis, 2002). An experience-based development process that includes experimentation and builds on professionalism of employees (whose performance is being measured) enhances both the validity and acceptance of the PMS.

The study shows, furthermore, that local PMS initiatives have a high chance of being successfully implemented, despite top-management's efforts to coerce the local unit into a much faster and consequently less-developmental mode. The action research described here stimulated learning, both on the part of the organization's department and by the university-based participants. After the study, there was a continuation of similar PMS activities, carried out by members of the logistics department with less active engagement of the university research partners. When reflecting on the project (in July 2006), members of the LMT emphasized the importance of management support for PMS development, because the development process requires significant time of employees at various levels, spent on activities that may not be seen as particularly "strategic" or "glamorous". Employees' individual expertise, insights, and skills, as well their enthusiasm for performance measurement, were utilized in the process; and they received the credits for results achieved, in terms of an enriched PMS. According to these LMT members, the development process had also benefited from the stimulating and challenging interaction with outsiders—researchers and students in this case.

Although the study is based on a multitude of observations, an obvious but important limitation is that it is based on a single case study. Results may be difficult to generalize to other empirical settings, also because the researchers and research assistants have not only been neutral observers; they were also involved in helping to expand and refine the departmental PMS. However, against these weaknesses stands the advantage that detailed observations could be made. Discussions with members of the organization where always lively, detailed, and involved. Our ideas were critically challenged, because ideas pertained to "their" PMS and the development and actual usage of it. These interactions were not discussions about abstract ideas in the interest of the researchers' project or theory. Rather they dealt with what made sense to organizational members in the context and language of their own work. We feel that our study, while acknowledging the limitations in terms of possible biases (for example, selective perception and interpretation), validly captures the departmental members' attitude toward performance measurement and the development process that had contributed to that attitude. Furthermore, the insights based on qualitative data have been complemented with representative quantitative data gathered within the department at two points in time.

An important field for future research remains the dualistic role of performance-measurement systems—to provide some of the knowledge necessary for planning and decision making, but also to motivate and monitor people in organizations (Zimmerman, 1997, p. 5)—and the effects of the incompleteness of such systems. We concur with Ahrens and Chapman (2004, p. 298) that "the concept of enabling control presents a clearly defined framework within which future research ... might further develop our understanding of the ways in which management control systems can simultaneously support the objectives of efficiency and flexibility". Future research could possibly expand our focus to conditions when a developmental, enabling PMS approach is most feasible. Feasibility is an issue, because this approach is demanding on employees, senior management, and support functions. A developmental approach assumes informal local experience with quantitative performance measurement, and that employees are quite willing and capable to build on that in order to further develop the PMS. Beyond that, we expect that other requirements also play a role. For example, time is needed to really understand in detail what is already in place, and to evaluate what will be reused and what not. Time and local autonomy are needed to not "fix" the PMS too soon so that improvements and adjustments to local conditions can be made. Furthermore, senior management needs to have a clear understanding of their objective for developing a PMS: is it to monitor and report upward in the hierarchy, or is it also (or even primarily) intended to support lower-level employees in their work? Senior management also needs to behave in accordance with an enabling PMS: balancing between recognizing the incompleteness of the PMS (so there is a story next to measured outcomes) and demanding certain performance. And a developmental approach needs to be facilitated in terms of resources and rewards, such as time to work on it, bestowing prestige upon PMS developers, support from experts, development of IT tools with which non-specialists can work, etc. Such facilitation requires high-level support from IT, and cross-functional cooperation of finance and accounting professionals. In sum, the developmental approach reported in this paper is demanding and may not be *feasible* in every organization.

The developmental approach may also not be equally *relevant* to every organization. While incompleteness of PMS helps to understand that a developmental approach affects the enabling nature of PMS, in certain organizations PMS may be well developed and stable. A developmental approach to shape the PMS may also seem less relevant if operations managers have other kinds of information that are more informative than formal performance measures, such as direct observations of processes. So, future studies may focus on the question: what are antecedents of an effective developmental PMS approach? Furthermore, investigating the benefits to the organization (such as employee learning, or financial benefits), as well as assessing other consequences of a developmental approach is an intriguing line of future research.

To conclude, this study analyzed and illustrated a developmental approach to PMS development, which harvests existing informal measurement practices, it lets new measurement experiments blossom, and at times prunes the extant measurement system. This developmental approach works through employees' local measurement experiences, experimentation with refined and new measures, and mobilizes employees' professionalism. Future research could help to better understand antecedents and consequences of this developmental approach toward performance-measurement systems.

Acknowledgements

The authors thank Thomas Ahrens, Chris Chapman, Kim Langfield-Smith, Jeff Hicks, the journal's Editor and Reviewers, company employees, and workshop participants at the Academy of Management 2005, the Warwick Business School, the Global Management Accounting Research Symposium 2006, and the New Directions in Management Accounting Conference 2006 for their comments and suggestions.

Appendix

This appendix contains the questionnaire items for the newly developed constructs *Attitude toward performance measures* and *Professionalism.*

Attitude toward performance measures

"In your department a number of performance measures" (or KPIs: "Key Performance Indicators") are used, as shown in the appendix to this questionnaire. We ask your opinion about the KPIs within your department. Please give a score from 1 to 7.

1. How familiar are you with the KPIs of your department?
2. How understandable do you find the KPIs of your department?
3. How reliable do you consider the KPIs of your department?
4. How validly reflect the KPIs the performance of your department?
5. How extensively are the measurements of the KPIs used in your department?
6. How involved are you within your department in the development of better KPIs?
7. How do you experience the process of developing better KPIs within your department?

8. How useful do you consider the present departmental KPIs for the Logistics Department?
9. How useful do you consider the present departmental KPIs for your department?
10. How useful do you consider the present departmental KPIs for you "personally?"

These questions were answered on seven-point Likert scales anchored to the key concept in each question. For example, "How reliable do you consider the KPIs of your department?" was anchored on "very unreliable", "unreliable", "somewhat unreliable", "neutral", "somewhat reliable", "reliable", "very reliable".

Professionalism

"We ask your opinion about the following statements". You can indicate the extent to which you agree with each statement by a number from 1 through 7:

Very much disagree	Disagree	Moderately disagree	Neutral	Moderately agree	Agree	Very much agree
1	2	3	4	5	6	7

1. I always contribute to new ideas at work.
2. At work, I like to be active improving things.
3. I like to do things well in my work.
4. The way I conduct my activities, is very consistent with what is being recommended by professionals.
5. I obey the rules at work.
6. I adhere to standards of integrity that pertain to my work.
7. I sometimes act in ways I should not, because it will not be noticed anyway (reversely coded).
8. The manner of my daily work I consider "professional".
9. The way in which may work is organized is professional.
10. I am busy with my profession or work also outside working hours.
11. I can demonstrate to other people that my work is important.
12. I learn every day at work.
13. I have colleagues at work from whom I learn.
14. I enjoy reading about my profession or work.
15. I take part in activities outside working hours that improve my professionalism.
16. I am always keen to follow suitable external courses.
17. I am always keen to follow suitable internal courses.
18. I learn from problems I encounter at work.
19. I am an active member of an organization (or network) that helps advancing my profession.
20. I get sufficient autonomy to direct my work.
21. I take my personal professional development seriously.
22. I keep myself informed about new developments in my profession or work.
23. I am actively improving at work.[4]
24. I would like to pursue more external training.[4]
25. I would like to pursue more internal training.[4]

References

Abernathy, M. A., & Brownell, P. (1999). The role of budgets in organizations facing strategic change: an exploratory study. *Accounting, Organizations and Society, 24*(3), 189–204.

Abernathy, M. A., & Lillis, A. M. (1995). The impact of manufacturing flexibility on management control system design. *Accounting, Organizations and Society, 20*(4), 241–258.

Abrahamson, E. (2000). Change without pain. *Harvard Business Review, 78*(4), 75–79.

[4] These items were added only after the first administration of the questionnaire.

Adler, P. S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly, 41*(March), 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. *Contemporary Accounting Research, 21*(2), 271–301.

Ahrens, T. A., & Chapman, C. S. (2006). Doing qualitative field research in management accounting: Positioning data to contribute to theory. *Accounting, Organizations and Society, 31*, 819–841.

Atkinson, A. A., & Shaffir, W. (1998). Standards for field research in management accounting. *Journal of Management Accounting Research, 10*, 41–68.

Baer, M., & Frese, M. (2003). Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior, 24*, 45–68.

Baines, A., & Langfield-Smith, K. (2003). Antecedents to management accounting change: A structural equation approach. *Accounting, Organizations and Society, 28*, 675–698.

Banker, R. D., Potter, G., & Schroeder, R. G. (1993). Reporting manufacturing performance measures to workers: An empirical study. *Journal of Management Accounting Research, 5*, 33–55.

Bass, B. M., & Avolio, B. J. (2000). *MLQ multifactor leadership questionnaire* (2nd ed.). Redwood City: Mind Garden (Technical Report).

Beamon, B. M. (1999). Measuring supply chain performance. *International Journal of Operations & Production Management, 19*(3), 275–292.

Bisbe, J., & Otley, D. (2004). The effects of the interactive use of management control systems on product innovation. *Accounting, Organizations and Society, 29*, 709–737.

Bourne, M., Neely, A., Mills, J., & Platts, K. (2003a). Implementing performance measurement systems: A literature review. *International Journal of Business Performance Management, 5*(1), 1–24.

Bourne, M., Neely, A., Mills, J., & Platts, K. (2003b). Why some performance measurement initiatives fail: Lessons from the change management literature. *International Journal of Business Performance Management, 5*(2–3), 245–269.

Caldwell, S. D., Herold, D. M., & Fedor, D. B. (2004). Toward and understanding of the relationship among organizational change, individual differences, and changes in person-environment fit: A cross-level study. *Journal of Applied Psychology, 89*(5), 868–882.

Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science, 13*(4), 442–455.

Carmona, S., & Grönlund, A. (2003). Measures vs actions: The balanced scorecard in Swedish law enforcement. *International Journal of Operations & Production Management, 23*(11), 1475–1496.

Cavalluzzo, K. S., & Ittner, C. D. (2004). Implementing performance measurement innovations: Evidence from government. *Accounting, Organizations and Society, 29*, 243–267.

Chapman, C. S. (1997). Reflections on a contingent view of accounting. *Accounting, Organizations and Society, 22*, 189–205.

Chow, C. W., Kato, Y., & Merchant, K. A. (1996). The use of organizational controls and their effects on data manipulation and management myopia: A Japan vs U.S. comparison. *Accounting, Organizations and Society, 21*(2–3), 175–192.

Davila, A., & Wouters, M. (2005). Managing budget emphasis through the explicit design of conditional budgetary slack. *Accounting, Organizations and Society, 30*, 587–608.

de Haas, M., & Algera, J. A. (2002). Demonstrating the effect of the strategic dialogue: Participation in designing the management control system. *Management Accounting Research, 13*, 41–69.

Eccles, R. G. (1991). The performance measurement manifesto. *Harvard Business Review, 69*(1), 131–137.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*, 350–383.

Fullerton, R. R., & McWatters, C. S. (2002). The role of performance measures and incentive systems in relation to the degree of JIT implementation. *Accounting, Organizations and Society, 27*, 711–735.

Hall, R. H. (1968). Professionalization and bureaucratization. *American Sociological Review, 33*(10), 92–104.

Hall, R. W., Johnson, H. T., & Turney, P. B. B. (1990). *Measuring up: Charting pathways to manufacturing excellence*. Homewood, IL: Business One Irwin.

Jaworski, B. J., & Young, S. M. (1992). Dysfunctional behavior and management control: An empirical study of marketing managers. *Accounting, Organizations and Society, 17*(1), 17–35.

Jönsson, S., & Grönlund, A. (1988). Life with a sub-contractor: New technology and management accounting. *Accounting, Organizations and Society, 13*(5), 512–532.

Kaplan, R. S. (1983). Measuring manufacturing performance: A new challenge for managerial accounting research. *The Accounting Review, 58*(4), 686–705.

Kaplan, R. S. (Ed.). (1990). *Measures for manufacturing excellence*. Boston, MA: Harvard Business School Press.

Labro, E., & Tuomela, T.-S. (2003). On bringing more action into management accounting research: Process considerations based on two constructive case studies. *European Accounting Review, 12*(3), 409–442.

Lillis, A. M. (2002). Managing multiple dimensions of manufacturing performance—An exploratory study. *Accounting, Organizations and Society, 27*, 497–529.

Lowe, A., & Jones, A. (2004). Emergent strategy and the measurement of performance: The formulation of performance indicators at the microlevel. *Organization Studies, 25*(8), 1313–1337.

Maiga, A. S., & Jacobs, F. A. (2005). Antecedents and consequences of quality performance. *Behavioral Research in Accounting, 17*, 111–131.

Malina, M. A., & Selto, F. H. (2001). Communicating and controlling strategy: An empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research, 13*, 47–90.

Maskell, B. H. (1991). *Performance measurement for world class manufacturing—A model for American companies*. Cambridge, MA: Productivity Press.

McKinnon, S. M., & Bruns, W. J. (1992). *The information mosaic*. Boston, MA: Harvard Business School Press.

Medori, D., & Steeple, D. (2000). A framework for auditing and enhancing performance measurement systems. *International Journal of Operations & Production Management, 20*(5), 520–533.

Nanni, A. J., Dixon, J. R., & Vollmann, T. E. (1992). Integrated performance measurement: Management accounting to support the new manufacturing realities. *Journal of Management Accounting Research, 4*, 1–19.

Otley, D. (1999). Performance management: A framework for management control systems research. *Management Accounting Research, 10*, 363–382.

Perera, S., Harrison, G., & Poole, M. (1997). Customer-focused manufacturing strategy and the use of operations-based non-financial performance measures: A research note. *Accounting, Organizations and Society, 22*, 557–572.

Piderit, S. K. (2000). Rethinking resistance and recognizing ambivalence: A multidimensional view of attitudes toward an organizational change. *Academy of Management Review, 25*(4), 783–794.

Qu, S. Q. (2006). *Translating popular accounting ideas into action: The role of inscriptions in customizing the balanced scorecard*. University of Alberta, School of Business.

Ramaswami, S. N. (1996). Marketing controls and dysfunctional employee behaviors: A test of traditional and contingency theory postulates. *Journal of Marketing, 60*(2), 105–120.

Ramaswami, S. N. (2002). Influence of control systems on opportunistic behaviors of salespeople: A test of gender differences. *Journal of Personal Selling & Sales Management, 22*(3), 173–188.

Reagans, R., Argote, L., & Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science, 51*(6), 869–881.

Scott, T. W., & Tiessen, P. (1999). Performance measurement and managerial teams. *Accounting, Organizations and Society, 24*, 263–285.

Simons, R. (1990). The role of management control systems in creating competitive advantage: New perspectives. *Accounting, Organizations and Society, 15*, 127–143.

Simons, R. (1991). Strategic orientation and top management attention to control systems. *Strategic Management Journal, 12*(1), 49–62.

Simons, R. (1994). How new top managers use control systems as levers of strategic renewal. *Strategic Management Journal, 15*(3), 169–189.

Simons, R. (1995). Control in an age of empowerment. *Harvard Business Review, 73*(2), 80–88.

Stanton, J. M., Balzer, W. K., Smith, P. C., Parra, L. F., & Ironson, G. (2001). A general measure of work stress: The stress in general scale. *Educational and Psychological Measurement, 61*(5), 866–888.

Snizek, W. (1972). Hall's Professionalism scale: An empirical reassessment. *American Sociological Review, 37*(1), 109–114.

Sorge, A., & Witteloostuijn, A. van. (2004). The (non)sense of organizational change: An essay about universal management hypes, sick consultancy metaphors, and healthy organization theories. *Organization Studies, 25*(7), 1205–1231.

Swailes, S. (2003). Professionalism: Evolution and measurement. *The Service Industries Journal, 23*(2), 130–149.

Townley, B., Cooper, D. J., & Oakes, L. (2003). Performance measures and the rationalization of organizations. *Organization Studies, 24*(7), 1045–1071.

Van der Stede, W. A. (2000). The relationship between two consequences of budgetary controls: Budgetary slack creation and managerial short-term orientation. *Accounting, Organizations and Society, 25*, 609–622.

West, J., & Iansiti, M. (2003). Experience, experimentation, and the accumulation of knowledge: The evolution of R&D in the semiconductor industry. *Research Policy, 32*(5), 809–825.

Yanow, D. (2004). Translating local knowledge at organizational peripheries. *British Journal of Management, 15*, S9–S25.

Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science, 13*(3), 339–351.

Zimmerman, J. L. (1997). *Accounting for decision making and control* (2nd ed.). Chicago, IL: Irwin.

Using Prototypes to Induce Experimentation and Knowledge Integration in the Development of Enabling Accounting Information*

MARC WOUTERS, *University of Twente*

DIANA ROIJMANS, *University of Twente*

1. Introduction

Performance measurement systems (PMS) serve different functions. They are formal devices for control, strategy formulation, and communication and as such primarily serve higher-level managers (Ittner and Larcker 2003; Kaplan and Norton 2006; Malina, Nørreklit, and Selto 2007; Simons 1991, 1995). However, we assert that PMS should also support the people whose performance is being measured. We refer to "enabling PMS" as those perceived by employees as facilitating their responsibilities, rather than primarily as control devices for use by senior management (Adler and Borys 1996; Ahrens and Chapman 2004; Free 2007; Wouters and Wilderom 2008). We examine the informing roles accounting information may play and how managers close to the shop floor can make more effective use of it in dealing with operational issues. Rather than viewing accounting as a formal control mechanism for senior management, derived from large information systems, or developed using organization-wide initiatives, we investigate small-scale, local, enabling accounting information. There is scope for different types of accounting information (Vaivio 2006), and an organization may benefit from accounting that allows experimentation, pluralism, and the presence of multiple information channels and systems (Hedberg and Jönsson 1978).

Previous studies in accounting have identified *system* characteristics of enabling PMS (Ahrens and Chapman 2004; Free 2007) and characteristics of the *processes* used in the design and implementation of enabling PMS (Wouters and Wilderom 2008). The enabling nature of PMS is fostered by a development process that allows room for experimentation with new performance measures (Wouters and Wilderom 2008). Experimentation refers to the use of trial-and-error cycles that repeatedly generate and test design alternatives, where each trial generates new insights (Thomke 1998; Thomke, von Hippel, and Franke 1998). While previous research shows that experimentation is important for the effectiveness of firms' innovation processes (Thomke et al. 1998; Tidd and Bodley 2002; West and Iansiti 2003), its use has not been considered in the context of designing and implementing enabling PMS. Hence, the research question addressed by this study is: How can experimentation stimulate knowledge integration for the development of enabling PMS?

The notion of experimentation builds on — but goes beyond — previous studies in accounting that have examined variables influencing the adoption and implementation of

* Accepted by Alan Webb. We thank Chris Chapman, Thomas Ahrens, Sander van Triest, Robert Chenhall, Anne Lillis, Paolo Quattrone, Alexandra Van den Abbeele, Kari Lukka, Michael Habersam, Eli Pernot, Kristof Stouthuysen, Martine Cools, Stijn Masschelein, Sofie Verbieren, Bart Dierynck, the Associate Editor Alan Webb, the reviewers, and participants and discussants at the University of Amsterdam, the European Academy of Management Conference 2007, Said Business School, Katholieke Universiteit Leuven, Monash University, American Accounting Association MAS Midyear Meeting 2009, and the University of Innsbruck. We thank Celeste Wilderom and company employees for their comments and insights. We thank the company for participation in this research project.

doi:10.1111/j.1911-3846.2010.01055.x

accounting innovations (Anderson and Young 1999; Gosselin 2007; Krumwiede 1998). Allowing users to influence the design of the accounting innovation can positively affect organizational performance, because user participation may increase the extent to which the system is tailored to their needs and may decrease measurement error (Abernethy and Bouwens 2005). Development of PMS often involves many challenges in selecting and interpreting appropriate performance metrics (Cavalluzzo and Ittner 2004). By letting users influence the design of the accounting innovation, their relevant tacit knowledge can be utilized in the development process. Technical issues such as data limitations and difficulties defining appropriate measures appear to play a more important role in the implementation of PMS than they do in costing system implementations (Cavalluzzo and Ittner 2004). While there is empirical support for the importance of user participation in the design and implementation of accounting innovations (Chenhall and Euske 2007; McGowan and Klammer 1997), far less research has examined how such participation can be shaped.

User participation implies that accountants and nonaccountants work together and combine their knowledge in developing enabling PMS. *Knowledge integration* refers to the ability of organizations to build practices that draw on diverse bases of expertise and has been identified as a key organizational capability in various fields such as strategic management (Eisenhardt and Martin 2000), new product development (NPD; D'Adderio 2001; Leenders and Wierenga 2002; Roller, Eck, and Dalakakis 2004), information systems development (Mitchell 2006; Levina and Vaast 2005; Patnayakuni, Rai, and Tiwana 2007), and accounting (Anderson 1995). However, knowledge integration involving accountants and nonaccountants is particularly challenging (Anderson 1995; Chapman 1998; Chenhall and Langfield-Smith 1998; Emsley 2005; Fry, Steele, and Saladin 1995; Sillince and Sykes 1995). For example, accountants and engineers tend to have quite different world views (Dent 1991; Ezzamel, Lilley, and Willmott 2004), and accountants are often hardly involved in the organization's engineering activities (Anderson and Sedatole 1998). Engineers think about products and operational processes in similar "technical" ways (e.g., shapes, physical inputs and outputs, functionality, performance, and time dependency between activities) and are knowledgeable about different aspects thereof. Conversely, accountants' representations of products and processes are related to notions such as cost allocations, inventory valuation, revenue recognition, and matching of expenses. (See also Eraut 2007.) As a result, we explore the role of accountants in more detail, in a context where knowledge integration is essential but particularly challenging: the development of *enabling* PMS.

Research on knowledge integration has focused on NPD and information systems development (Levina and Vaast 2005; McDonough 2000; Nambisan and Wilemon 2000). Experimentation with prototypes can stimulate knowledge integration (D'Adderio 2001; Carlile 2004), and we draw on the related literature because insights from other fields can advance knowledge in accounting research (Davila and Oyon 2008). Our results suggest that experimentation with the help of prototypes is a novel and relevant concept worthy of introduction to the accounting literature.

This study uses a longitudinal case study, involving action research. We investigated the development of a PMS at the operational level of a medium-sized company in the beverage manufacturing industry, and we participated in the development process. We find that expertise and data from diverse functional areas needed to be integrated in developing an enabling performance measurement system for the transportation department. The challenge faced by the company was to bring together different employees' understanding of: (1) how particular operational processes were conducted and (2) how these processes were represented in various information systems (accounting systems and operational systems). We identify characteristics of conducting the experimentation that helped to achieve user participation and knowledge integration.

The remainder of this paper is structured as follows. Section 2 provides our literature review and a more detailed motivation for the present study. Propositions about effective experimentation are developed in section 3. The research method is described in section 4. In section 5 the development of the PMS is described, and these findings are discussed in section 6.

2. Literature review

Enabling PMS and the role of experimentation

The framework of Adler and Borys 1996 has recently been introduced in the accounting literature by Ahrens and Chapman 2004. Employees are more likely to have a positive attitude toward formalized systems such as PMS when it motivates or enables them to better perform their tasks, but will be more negative if it functions as a means by which management attempts to coerce their effort and compliance (Adler and Borys 1996). Enabling PMS support managers whose performance is being measured. Enabling PMS are diverse and in some sense informal. They may not necessarily be part of a large centralized system, but instead could be an Excel application developed and maintained by a local manager. However, enabling PMS are not completely detached from formal systems. For example, they may: (a) be initiated in response to central pressures and initiatives, (b) download data from formal systems, (c) use other official resources, and (d) provide inputs to formal meetings and procedures (Ahrens and Chapman 2004; Frow, Marginson, and Ogden 2005).

Several characteristics of PMS contribute to their enabling nature. Adler and Borys (1996) propose *internal and global transparency, flexibility,* and *repair* as key characteristics that contribute to the enabling nature of formalized systems. Internal transparency means that users understand the logic of a system's internal functioning and have information on the system's status. Global transparency refers to the intelligibility for employees of the broader system and the context within which they perform their work. Such a system offers employees an understanding of where their own tasks fit in the organization as a whole. Flexibility means that users can make controlling decisions after enabling systems have provided information. "Flexible systems encourage users to modify the interface and add functionality to suit their specific work demands" (Adler and Borys 1996: 74). Repair means that users can mend and improve the work process themselves rather than allowing breakdowns and other nonprogrammable events to force work processes to a halt. Recent case studies by Ahrens and Chapman 2004, Free 2007, and Chapman and Kihn 2009 provide empirical support for these characteristics. Also, based on survey data, Naranjo-Gil and Hartmann (2006) found that the coercive-enabling framework helped to explain the use of accounting information by top management teams.

The *process* through which a PMS is developed may also contribute to the enabling nature of the resulting system, but this has been investigated to a lesser extent. Otley notes, "The balanced scorecard literature also indicates that it [is] as much the process of establishing a scorecard that yields benefit as the resultant measurement schema. However, the literature is remarkably silent on this point" (1999: 377). Wouters and Wilderom (2008) explicitly focus on the process of developing enabling PMS, and they find that *experimentation* is a key characteristic. This is because enabling PMS require measures that are closely related to the specific operating conditions in a particular setting, that are actually measurable (i.e., the required data are available), and that are presented in a way that employees find understandable. The making of a performance measure is not likely to be "right" after just one attempt. Instead, it is more likely to be successful if the development engages employees in experimenting with conceptual ideas, data, and representations for the new measures. Hence, experimentation involves first the development of a performance measure and then subsequent testing and refinement (in several rounds) of its conceptualization, definition, required data, information technology (IT) tools, and presentation. This will involve employees whose performance is going to be measured to arrive at information that is a valid, reliable, and understandable indicator of performance in a specific context. This is a meticulous and creative process, and the performance measures under construction that are generated and reviewed constitute prototypes.

Other studies in accounting also provide support that experimentation may contribute to the enabling nature of the resulting PMS. Abernethy and Bouwens (2005) find that managers accept accounting innovations better when they can influence the design of the accounting system. The managers' involvement enhances the extent to which the system is tailored to their needs and increases the validity of performance measurement. Malina et al. (2007) find that a system that motivates and that is seen as legitimate and fair is often based on learning by doing and on continual improvement of system reliability. Tuomela (2005) finds that developing PMS is likely to require an evolutionary process, during which such a system can be gradually refined. More generally, user participation is important for successful implementation of management accounting systems (McGowan and Klammer 1997).

But experimentation is not a panacea. First, a process of heavy experimentation is demanding on employees, senior management, and support functions (Chenhall and Euske 2007). Time and local autonomy are needed that allow some flexibility for adjusting and improving the PMS. Employees need to have time for this, a sufficient level of professionalism, and other resources (Wouters and Wilderom 2008). Second, trust between employees and managers is required, so that employees are willing to reveal private knowledge about how their processes can be made more transparent, instead of using the process to create slack and hide unfavorable results. Managers must also respect suggestions made by employees and must not abuse the system later for control purposes. Third, the notion of an enabling PMS assumes that managers intend the system to support the work of employees, but many PMS are implemented primarily for control purposes (Townley, Cooper, and Oakes 2003). So, the experimental approach is not without costs and may not be feasible in every organization.

Unfortunately, there is very little research examining the effects of user participation in developing accounting information (Abernethy and Bouwens 2005). There is considerable research examining the consequences of budgetary participation (the amount of influence a subordinate manager has for setting her unit's budgets). A recent meta-analysis based on 59 independent samples (Derfuss 2009) found 11 relationships between participative budgeting and its consequences that are homogeneous and generalize across samples (e.g., the positive effect of budgetary participation on budget usefulness). However, results also show relationships that are heterogeneous across samples (e.g., the equivocal effect on managerial performance).

In sum, there are indications that experimentation may contribute to the enabling nature of PMS, although this may not be a preferred or feasible approach in every organization. More generally, there is also empirical support for user participation in accounting. However, the literature has not addressed *how* users can be engaged in the experimentation process.

Knowledge integration for developing PMS

Knowledge integration is vital for the development of enabling PMS, because such a system needs to represent knowledge from many different people, such as specialists from finance, accounting, and IT, as well as users and managers who will use the system. Such cross-functional cooperation is important for developing PMS that are tailored to the management practices in specific situations (Anderson 1995; Chenhall 2008; Chenhall and Euske 2007; Chenhall and Langfield-Smith 1998; Emsley 2005; Fry et al. 1995; Sillince and Sykes 1995; Vaivio 2004). Several studies have demonstrated that nonaccounting ownership contributes to adoption of management accounting innovations (Shields 1995; Wouters, Wynstra, and Anderson 2005).

We argue that the effective development of enabling PMS requires an understanding of how people monitor processes and results, solve problems, take action, and influence costs, as well as how these management actions can be reflected in the detailed conceptualization of performance measures (Ahrens and Chapman 2002; Vaivio 2006). Local managers observe and intervene in operational processes, and they utilize different resources to achieve diverse and potentially conflicting goals. They combine information from various sources (such as observations, conversations, e-mails, phone calls, meetings, nonfinancial information systems, and formal reports) to monitor the current status of processes, to predict whether goals will be achieved, and to respond to setbacks and other unexpected situations (Jönsson and Grönlund 1988; McKinnon and Bruns 1992). The usefulness of accounting information at the operational level depends on how well it addresses the needs of this wider set of cost-management practices.

Several studies demonstrate that accounting information — in order to be useful — needs to be consistent with how operational managers influence costs and performance. For example, Van der Veeken and Wouters (2002) show in their case study of cost management by site managers in a road building company that these managers were closely involved in the planning and budgeting of a project. As a result, they formulated observable milestones to monitor actual costs and progress during project execution. If they observed that actual progress and costs were not meeting these milestones (and the budget) they employed a variety of practices to expedite the project, reduce costs, or increase revenues. Identifying these management practices was key to understanding why planning and budgeting were crucial activities for these site managers. Explicating these practices was also key for understanding why the managers made limited use of a new information system for monitoring actual project costs during project execution. Instead, they used other, more practical sources of information for assessing the status of operations and for solving cost management problems. A case study of a restaurant chain by Ahrens and Chapman (2002, 2004) also analyzes the role of accounting information at the level of operations management. The study first investigates how restaurant managers thought about their roles, what actions they took to achieve results, and how accounting information and systems supported them, if at all. Collectively, these studies demonstrate that it is challenging to understand what accounting means for different people and how it is used outside the finance and accounting domain, without even entering more complex settings with greater uncertainty (Chapman 1998; Chenhall and Morris 1995; Jørgensen and Messner 2010).

Enabling PMS in operations are highly specific to the operational characteristics and ways in which local managers can monitor and influence costs. Hence, the development of such systems requires knowledge integration involving employees for whom the enabling PMS is intended as well as specialists from finance, accounting, and IT. However, cross-functional cooperation between accountants and operational managers can be problematic (Chenhall and Langfield-Smith 1998). In addition, in a review of the literature on cross-functional teams and other types of horizontal organizations, Chenhall (2008) finds that management accountants are not seen as particularly relevant. Also, Rowe (2004) finds that, despite the pervasiveness of cross-functional collaboration in organizations, this is an underexplored area of accounting.

Knowledge integration has been studied in the context of NPD more than it has in accounting (Leenders and Wierenga 2002; McDonough 2000; Song, Thieme, and Xie

1998; Souder, Sherman, and Davies-Cooper 1998). An advantage of cross-functional teams is the resultant improvement of horizontal communication linkages (Souder et al. 1998; Galbraith 1973), and the variety of perspectives in a team may stimulate innovation and increase the likelihood of producing more successful new products (Song et al. 1998). Experimentation with prototypes stimulates cross-functional knowledge integration, because creating prototypes explicates functional knowledge in different domains, and the prototype can be the focal point for evaluation, discussion, and innovation from different functional perspectives (D'Adderio 2001). Prototypes "have proved effective in providing a concrete means of representing different functional interests and facilitating their negotiation and transformation in product-development settings" (Carlile 2004: 559). Such integrating devices may take several forms, for example, sketches, realistic physical prototypes, scale models, three-dimensional computer-aided design pictures, maps, Gantt charts, or workflow simulations (Bechky 2003; D'Adderio 2001; Thomke 1998). Research on information systems development has also found that a prototype provides the user with a tangible means of comprehending and evaluating the proposed system (O'Leary 1988; Eva 2001; Gordon and Bieman 1994; Davis and Venkatesh 2004; Hunton and Beeler 1997). However, in the context of new accounting information development, we know far less about cross-functional knowledge integration and how prototypes might be helpful.

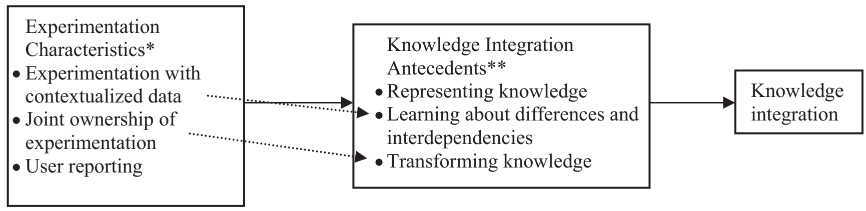
To conclude, the development of enabling accounting information requires an understanding of what actions managers take to influence costs and how these actions can be supported by accounting and other types of information. This requires knowledge integration, because knowledge about operational processes, cost management practices, and information systems is dispersed across different people. Knowledge integration is difficult, and the NPD literature suggests that it may be facilitated by experimentation with prototypes (Carlile 2004; D'Adderio 2001). However, the use of prototypes as a means of facilitating knowledge integration has not been investigated in an accounting context.

3. Proposed characteristics of experimentation for knowledge integration

This section develops a number of characteristics of experimentation for effective knowledge integration in the development of enabling PMS. We present these ideas before the results of the empirical study because this structure helps to better clarify our theoretical ideas in relation to the literature, as well as the empirical findings in relation to these ideas. This structure should not suggest that these ideas have been developed before doing the empirical study. Rather, our ideas about effective experimentation were developed by going back and forth between the literature and the observations, while still being active at the case study company (Ahrens and Chapman 2006).

Knowledge integration is particularly important when communication is needed across functional boundaries within organizations, and individuals need to work across such boundaries. According to Carlile, "what we see at a pragmatic knowledge boundary is not just a matter of processing more knowledge, but processes for transforming knowledge" (2002: 453). Transforming knowledge "refers to a process of altering current knowledge, creating new knowledge, and validating it within each function and collectively across functions" (445). Prototypes and other integrating devices can support knowledge integration across knowledge boundaries, and Carlile (2002) proposes three characteristics of effective integrating devices (see also Figure 1): (1) *representing:* prototypes establish a shared syntax or language for individuals to represent their knowledge, (2) *learning:* prototypes provide concrete means for individuals to specify and learn about their differences and interdependencies across a knowledge boundary, and (3) *transforming:* prototypes facilitate a process where individuals can jointly transform their knowledge. For example, a physical prototype of the engine compartment of a car represents the knowledge from different design teams about the three-dimensional shapes of the various systems that need to find a place inside

Figure 1 Characteristics of effective experimentation for knowledge integration, in the context of developing enabling performance measurement systems.



Notes:

*Characteristics identified in this study that enhance the effectiveness of the knowledge integration antecedents. The dotted arrows reflect more specific relationships: Experimentation with contextualized data especially enhances Learning about differences and interdependencies; Joint ownership of experimentation particularly augments Transforming knowledge.

**Characteristics of integrating devices identified by Carlile 2002 as key antecedents to knowledge integration.

that compartment, such as the engine, transmission, climate system, braking system, and suspension. Interdependencies and conflicts become clear, such as systems not fitting in the same space, assembly being inefficient, or certain repair activities being too complicated; using a prototype can facilitate finding and negotiating a solution. These three characteristics point to both the practical capacity and political capacity of prototypes as integrating devices — developing an adequate common knowledge is also a political process of negotiating and defining common interests (Carlile 2002, 2004).

These three characteristics will be discussed in more detail with respect to knowledge integration in developing PMS. First, in the context of experimenting with new accounting information, prototypes of new performance measures represent knowledge that different people have brought to the process. For example, they will have ideas about what the content of the new information should be and how this information should be presented; gradually the prototype starts to represent these ideas. Or, the people involved may have different ideas about specific definitions of performance measures and data available from particular information systems that are consistent with these definitions; the prototype can start to include such information. Second, because experimentation around a joint prototype is being conducted, people will learn about each other's ideas — not just in general terms but as specifically operationalized in the tangible prototype. As Chenhall and Euske (2007) find in their study, accounting systems can be a channel for bringing together information from many sources: "The systems sought to provide a single data base with a common language" (634). Third, because the different participants involved are probably not always in agreement about the new performance measures, the process of experimenting with prototypes may help them to discover and agree on ways to resolve such differences; participants are transforming their knowledge.

We propose three further characteristics of effective experimentation in the development of enabling PMS. These characteristics reinforce the sharing of knowledge, highlight the discovery of differences and interdependencies between accountants and other professionals and managers, and strengthen the necessity to solve problems by negotiating solutions. As a consequence of these factors, representing, learning, and transforming knowledge across pragmatic boundaries all increase, as represented in Figure 1. The three additional characteristics are as follows:

(1) *Experimentation with contextualized data*. Using real data taken from real information systems as input *during* experimentation is a key characteristic to stimulate representing, learning, and transforming knowledge. In other words, examples (prototypes) of new performance measures are constructed with data pertaining to earlier periods that are downloaded from enterprise resource planning (ERP) systems, accounting systems, operational databases, and other information systems on which the new PMS will eventually draw. This use of contextualized data does not pertain to testing an information system using a large data set after its development (e.g., to investigate system performance). Instead, it means that users already use actual data and information systems while conceptualizing the PMS and experimenting with these conceptualizations. This is more demanding than just devising a fictitious numerical example of new performance measures. However, experimentation in this way is likely to increase knowledge integration because such experimentation can reveal more existing differences and interdependencies. For example, employees may realize the dissimilarity of their ideas about the appropriate definitions and data for specific performance measures or the controller's office finds out that for maintaining the new performance measures particular data must be provided by operational users. When prototyping and experimentation do *not* incorporate actual data from the company's information systems, there will be fewer opportunities to discover which knowledge is required, to explain that knowledge to other people, to realize differences in understanding, and to resolve them.

(2) *Joint ownership of experimentation*. Another stimulus for effective knowledge integration is that accountants, nonaccounting professionals, and managers jointly develop the new accounting information and know that they are required to reach agreement on the PMS. During the development process there is experimentation with different versions of the PMS, and at some point decisions have to be made about the final design. Consequently, the question may arise as to who has the final authority for making the decision. We suggest that effective knowledge integration requires that both accounting and operations participants agree on the design. Neither of these groups should make the final design decision independently. The requirement that they should jointly agree on the design of the system fosters (or forces) efforts to gain a better understanding of what the other parties find particularly important, why that is the case, and how any differences can be resolved. We propose that experimentation is more effective as an enabler of cross-functional knowledge integration if accounting and operations personnel realize at the beginning of the process that agreement is going to be required. This joint ownership of the development process requires negotiation when different interpretations and preferences for the accounting information arise, thus promoting knowledge integration. Such joint experimentation is likely to reinforce representing, learning, and transforming knowledge, the characteristics discussed above based on Carlile 2002, 2004.

(3) *User reporting*. We propose, furthermore, that effective experimentation benefits from an arrangement where the operational user has a significant role in actually producing the periodic performance report after its implementation. The user may either completely prepare such a report or simply provide key data to the controller's office. This means that the operational user needs to understand not only the meaning of the output, but also details as to which inputs are required and how the output is generated. As a result, users may realize that producing the reports will require some time and effort (compared to automatically receiving reports from the accounting department), and this may motivate them to invest more into making the PMS as relevant as possible. We propose that such involvement and mutual dependency are likely to increase

the level of required knowledge integration, because the interdependencies are enhanced.

In this section we have proposed a number of characteristics of experimentation with new performance measures. It is a demanding process, as mentioned previously, and joint ownership of experimentation may increase the lead time required to put new performance measures in place, and user reporting may be difficult because of the required time and commitment of employees and the temptation to manage the results. But the reason for going through an experimental process using prototypes lies in its potential to stimulate knowledge integration and to develop enabling accounting information. The empirical results, presented in the next section, will provide a further understanding of and support for these propositions.

4. Research method

Because few studies in accounting have addressed our research question, this research is exploratory, based on a case study of the logistics department of a beverage manufacturing company. The researchers contributed to the development of new accounting information over several years through longitudinal action research, and the specific study reported here spans a 10-month period (September 2005 to July 2006). We will first introduce the research site and then the motivation for longitudinal action research.

Research site

The case study organization operates in the beverage manufacturing industry, with sales of around 300 million euros in 2005. Within the company's Logistics department, one of the departments is Physical Distribution, and within that is the department of External Transportation, where this study was conducted. The company sells to the hospitality industry (bars, restaurants, and hotels) and to retail customers (mainly supermarkets). The External Transportation department is responsible for transportation activities between the manufacturing site and customer locations (such as distribution centers or individual bars and restaurants) or other outside locations. The study focuses on the cost of domestic transportation, excluding bulk transportation, and international transportation.

This study was part of a larger research effort with this company, which spanned about four years (August 2002 to July 2006) with ongoing informal contacts since then.[1] The research focused on the process of developing enabling PMS in the Logistics department. We interacted with more than 25 different employees from several departments across the organization, through dozens of various meetings, interviews with one or a few employees, participation in meetings with a larger number of employees, and informal off-site meetings (e.g., diners, sponsored hospitality events). Furthermore, we made phone calls and exchanged e-mails with company employees. Discussions with employees during meetings and phone calls were recorded by taking notes, which we worked up directly afterwards to capture more details from memory. In some of the meetings both researchers were present. Furthermore, we obtained a large number of sample documents, employees completed several rounds of a survey, and several research assistants worked full time at the company for periods of between six and 10 months. Table 1 provides more details.

For this particular study, one of the researchers worked full time at the company, first as an intern to write a master's thesis on this project and subsequently (since June 2006) as a company employee. Coghlan (2001) describes this process as "insider action

1. For example, a few MSc students did internships at the company, we spoke with company officials about ongoing developments in performance measurement, and during February 2008–March 2009 another study took place in a different department.

TABLE 1
Data collection—August 2002 to July 2006[1]

Employee meetings:[2]	Number	Total duration (hours)
Logistics only	11	20
Outside Logistics	20	25
Logistics together with other functional areas	25	37
Total	56	82
Different employees interacted with:	**Number**	
Logistics	9	
Finance	7	
Production	2	
Marketing and Sales	2	
Other functional areas	8	
Total	28	
Sample company documents:[3]		**Number**
Documents about performance measures used by Logistics		21
Documents about performance measures used outside Logistics		4
Presentations and documents about performance measurement at the company		8
Minutes of meetings about performance measurement at the company		18
General documents about the Logistics department		11
General company documents		9
Total		71

Notes:

1 The results reported in this paper pertain to the period September 2005 to July 2006, which focuses on the concept of experimentation. The importance of experimentation for developing enabling PMS became clear in an earlier phase of the research program (Wouters and Wilderom 2008).

2 "Meetings" indicates face-to-face meetings between the researchers and members of the case-study organization, either as interviews with one or a few employees or as active participation in meetings with a larger number of employees. Most meetings took place at the research site, with a few held at the researchers' university. In some cases we met with an employee more than once. Not included in these counts are e-mails, phone calls, and interactions between company employees and the research assistants.

3 These numbers include only documents obtained by the researchers (not the research assistants).

research". The other researcher was based at a university and was also actively involved in team meetings and provided expert input (unpaid) to the process. There was frequent interaction (often daily and at least weekly contact) with 10 different employees from different departments in the company. Also, a multidisciplinary project team (including both authors) of six individuals met about once every six weeks, there were numerous meetings with other employees throughout the project, and the researchers worked with the company's information systems and documents.

Longitudinal action research

The objective of longitudinal action research is to generate theoretical insights, but helping an organization to innovate accounting information can be a useful *means* to that end (Kasanen, Lukka, and Siitonen 1993; Jönsson and Lukka 2007). Understanding accounting change in an organization can take considerable time (Ahrens and Chapman 2006; Otley and Berry 1994), so we chose to study the development of enabling PMS over an extended period. Interactions over a longer time period create more familiarity with people in the organization and a better understanding of the organizational situation, which facilitate "repeated trials for approximating and understanding a research question or topic" (Van de Ven and Johnson 2006: 813).

Potential drawbacks of action research are that the researcher may unduly influence what happens in the organization and may selectively look for empirical evidence and guide the research process with a bias toward the expected findings (Atkinson and Shaffir 1998). However, there are several countervailing effects that limit such potential bias, which have been described recently in both the management accounting literature (Atkinson and Shaffir 1998; Labro and Tuomela 2003) and operations management literature (Coughlan and Coghlan 2002).

First, field research may promote construct validity because of the close observation of the organization (Atkinson and Shaffir 1998). "By observing people in their everyday lives, the field researcher obtains first-hand knowledge of social life that most accurately reflects the social world under investigation. By enabling us to stay close to the empirical world, qualitative methods yield a close fit between the data and what people actually say and do" (Atkinson and Shaffir 1998: 49). We were part of the developmental process, and we could directly observe discussions between team members, follow up with individuals, compare perspectives, and study documents and systems.

Second, action research may create a particularly in-depth interaction with the organization, compared to nonparticipant field research (Jönsson and Lukka 2007). The researchers are challenged to demonstrate to an organization that long-term research cooperation is likely to have practical relevance to the organization. The researchers are also challenged to innovate and construct a theoretically grounded solution, which can actually be implemented as part of the research process (Labro and Tuomela 2003). We found that people in the organization were very engaged. They expected results that were of practical relevance, and so there was an incentive for them to be involved and to spend time with the researchers. They were interested, challenged ideas, and provided feedback on results. It was not a linear process where the researchers designed frameworks and the organization implemented them. Instead, the process was a journey of joint discovery. As Van de Ven and Johnson (2006) note, research projects can be "collaborative achievements in learning among collaborating faculty, students, and practitioners" (811). This takes time, which "is critical for building relationships of trust, candor, and learning among researchers and practitioners" (812).

Third, it is crucial that field research starts with early ideas or hypotheses, however vague, that focus the initial observations. These ideas are then continuously revised and refined based on observations and new theoretical insights (Atkinson and Shaffir 1998). The early theoretical ideas for this particular study were explicitly documented in a research paper (Wouters and Wilderom 2008), which discusses notions about enabling PMS and the role of experimentation and other processes for developing such PMS. We then sought to further develop our understanding of experimentation, guided by both theory and empirical observations. During the development of the PMS in the company and also afterwards while writing this paper, the research process was characterized by iterating between data and theory, in what Dubois and Gadde 2002 label "systemic combining" (554).

Fourth, it is important that the research objectives be apparent to the organization and that the nature of the participation is clearly laid out when reporting on the study (Atkinson and Shaffir 1998). In this research project, there was an explicit agreement with the company that the efforts and cooperation would serve two purposes. Not only was it our intention to provide practically usable results, but the objective was also to gather empirical data for research purposes. The participation can be described as follows. The researcher in the company was the central person in the development process. She spoke with employees (individually and in groups) about their work, the conceptualization of the new accounting information, the available data, procedures, and information systems. She asked questions; helped brainstorm; and suggested ideas, listened to feedback, presented spreadsheets with performance measures, guided discussions, resolved questions and misunderstandings, and modified the prototypes in several cycles. Members of the organization have also read and discussed drafts of papers pertaining to this research project, spoken about the research in professional conferences, and coauthored papers for professional journals.

Action research as discussed here differs from how Kaplan (1998) describes working with companies in developing, implementing, and testing new accounting information (activity-based costing and the balanced scorecard). We did not set out to test or refine a particular accounting innovation; rather, our research objective was to theorize about the *development process*.

5. Developing the new accounting information

The company had several reasons for developing a PMS for transportation costs. In general, there was pressure on costs because of intensified competition, while at the same time the efficiency of transportation activities was jeopardized by more promotional sales. A large volume of the beverages was sold in bottles and crates for which consumers paid a deposit as an incentive to return these to the retail outlets after use, from where they were transported to the retailer's distribution centers. When a semi-trailer truck delivered new beverages to a retailer's distribution center, it also picked up the empty packaging materials for reuse. The beverages were often on promotion, which led consumers to buy inventory. As a result, some time before and during the promotional period, a higher volume was transported to the distribution centers of the retailer offering the promotion in its stores, without an equally higher return flow of packaging materials. After the promotional period, the volume sold to the retailer was lower, while consumers gradually returned the packaging materials to the stores. These imbalances increased transportation costs.

Another reason for developing a PMS was that the company had gradually switched from transportation using its own vehicles and drivers to full outsourcing of transportation. This shift changed the monitoring and budgeting of transportation costs. The company no longer had to manage the efficiency of the transportation activities because the transportation firm was responsible for executing transportation activities in accordance with standard driving times and standard distances. Still, the company continued to plan the transportation routes, so particular levers for managing transportation costs remained.

Third, limited information was available for understanding and managing costs. The transportation manager received monthly invoices from the transportation firm, the controller's office informed him about the actual numbers of pallets, and he produced a graph of the actual cost per pallet. He monitored the trend of actual costs, but an unexpectedly high or low cost was usually difficult to explain. The transportation manager's impression of how things had been going in a particular month could provide an understanding of why costs would be low, high, or average, but there was no analysis using accounting data. Both the operations departments and the controller's office felt the need to manage these costs more closely.

This section describes the development process for the new accounting information in three stages. First, the process focused on understanding the nature of transportation processes and imagining what new accounting information could be useful. Next it addressed what these initial ideas precisely meant and how the information could actually be produced. Finally it was concerned with institutionalizing the new information and developing reporting routines.

Forming initial ideas about the new accounting information: variance analysis

The first months centered on finding out what kind of information could be useful to better manage the costs of external transportation — "to better understand why we are above or below budget", as stated by one of the managers. The nature of transportation activities was discussed extensively. What did these activities entail, and what could the manager do to influence transportation costs? The information in this section represents knowledge that the transportation manager made explicit and shared with the others because of the joint development efforts. This information exchange led to the first ideas about variance analysis for transportation costs.

Standard driving times and standard distances for the planned routes were calculated by the transportation routing software, which had been developed by a consortium of transportation firms and companies purchasing transportation, and many companies used this software as the basis for invoicing. If the standard kilometers and hours differed from the actual kilometers and hours, the company did not bear the risk of such variances, at least not within the contracting period.

The following possibilities were identified as ways for the company to influence transportation costs: (1) Select the optimal types of trailers for each route. Some trailers used a special installation for loading and unloading, which moved all pallets on and off the trailer without the need for a forklift truck. Although this type of trailer could not be used for all customers, the planner should consider using this type of trailer whenever possible. (2) Use two-axle trucks when technically possible. (The special trailer mentioned above required a three-axle truck.) If such trucks were technically possible but not available, it would be more efficient to use a two-axle truck for the longer route and a three-axle truck for the shorter route.[2] (3) Plan routes so that the required resting of truck drivers would happen as much as possible during the unloading of trailers. (4) Combine transportation of empty packaging materials efficiently. Imbalanced flows sometimes led to additional efforts to pick up packaging materials (empty bottles and crates). In some cases special rides were necessary (e.g., going empty to a customer and driving back with crates and empty bottles), and sometimes a vehicle with some unused cargo space would take a detour to pick up such materials at another customer's site. The company could negotiate with customers when to pick up these materials, and the planner could efficiently incorporate such trips in the transportation planning.

Other ways to manage costs could be used over the longer term: (1) Influence the times needed for loading at the manufacturing site and through arrangements with customers influence the times needed for unloading at the customers' sites. While standard

2. The company used mainly semi-trailer trucks. This type of vehicle consists of a towing engine vehicle (also called truck) and a semi-trailer that carries the freight. The semi-trailer has wheels only in the rear, and the front is supported by the towing vehicle. The company used two different types of trucks (with two or three axles) and three different types of trailers, giving six different truck-trailer combinations. For brevity, henceforth we will use simply the term "truck" instead of "semi-trailer truck" to refer to a combination of truck plus trailer.

times for loading and unloading were invoiced (short-term deviations were the responsibility of the transportation firm), over time the company had succeeded in reducing these standard times. (2) Influence the utilization of vehicles through full truckload arrangements with customers. If customers ordered less than full truckloads, costs could be influenced through an efficient combination of orders leaving the manufacturing site in one trailer. (3) Make efficient arrangements with customers regarding delivery time windows and other conditions. These arrangements were not explicitly considered in the routing system. For example, it did not matter for determining the standard number of kilometers and hours whether a customer had a tight or otherwise inconvenient delivery time window that actually required building safety time into the schedule. However, in the longer term, more efficient arrangements could allow higher productivity and lower invoicing rates used by the transportation firm. (4) Include transportation costs in the assessment of customer profitability.

These various possibilities for influencing costs are related to the way transportation was contracted (standard costs for actual routes driven), and the options would differ if, for example, a fixed fee per transported pallet had been contracted.

The description above of the transportation activities and the possibilities for managing costs was the result of a detailed process of knowledge exchange. While participants understood the basic idea of trying to achieve the goal of having trucks and trailers fully utilized and return flows efficiently combined, the more intricate details described above became clearer through discussions with the transportation manager, especially early in the process. During those attempts to jointly conceptualize information that would be useful, in individual meetings and in team discussions, basic questions repeatedly came up, such as: "What affects the actual cost being higher or lower than the budget?"; "What do you do to manage costs?"; "What are the specific actions that you and others can take?" Team members tested each others' assumptions ("I thought that you . . ."), discussed ideas, and asked for clarification. The environment of developing new accounting information raised and legitimized these questions and discussions.

Gradually the idea of variance analysis emerged. It was clear that many different factors impacted actual costs, but it was concluded that trying to capture all factors would become too detailed. Also, many factors were not controllable by the company, but were the transportation firm's responsibility. The idea was to address costs at a more general level by separating the total difference between actual and budgeted costs into price and efficiency components, drawing on the traditional variance analysis approach in manufacturing. How the approach could be conceptualized in this setting, where the data would come from, and how it would be presented were not yet clear.

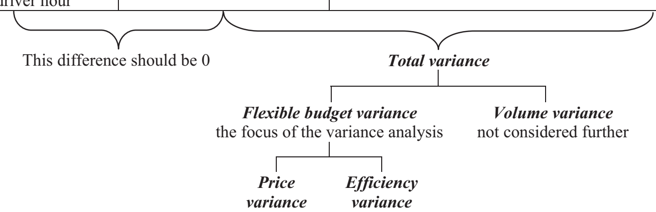
Designing and implementing the new accounting information

Subsequently, the process focused on developing the rough ideas about variance analysis further, and many details needed to be addressed. This process may appear to be just implementation, but that is not the case, as will be described. Working out the details was not part of implementation *after* design, but instead was actually part of the conceptualization and design of the new accounting information.

The final definition of the accounting information is described in Figure 2, and an example of one of the actual reporting graphs is shown in Figure 3. The difference between the original budget and the flexible budget was called the *volume variance*, which was not considered in further detail. The difference between *actual cost* and the *flexible budget* was called the *flexible budget variance*, and this was further disaggregated into price and efficiency variances. The *price variance* was calculated for each component separately (kilometers of trucks, hours of trucks, and drivers' hours) on the basis of the difference between the budgeted rate and the invoiced rate (e.g., standard rate per kilometer minus

Figure 2 Definition of the various cost systems used by the organization.

General ledger	Actual costs	Flexible budget	Budget
ERP system The transportation manager records invoices in the GL in the ERP system. The invoices specify total charges as per the contract structure shown below.	*New Excel worksheet* The calculation of actual costs is done on the basis of data from the routing system (plus rate information).	*Budgeting system* Budgeting is done in a spreadsheet by the controller's office. Data from the routing system are used to calculate operational parameters. *Route planning system* This system is used by the transportation manager and the transportation planner.	
Actual cost in the GL = $K_{i1}\,k_{r1} + K_{i2}\,k_{r2} + K_{i3}\,k_{r3} + K_{i4}\,k_{r4} + K_{i5}\,k_{r5} + K_{i6}\,k_{r6} + H_{i1}\,h_{r1} + H_{i2}\,h_{r2} + H_{i3}\,h_{r3} + H_{i4}\,h_{r4} + H_{i5}\,h_{r5} + H_{i6}\,h_{r6} + D_i\,d_r$	Actual cost (calculated) = $K_a\,k_{r,i} + H_a\,h_{r,i} + H_a\,(1\text{–}0.0625)\,d_{r,i}$	Flexible budget = $P_a\,(k_p\,k_r + h_p\,h_r + d_p\,d_r)$	Budget = $P_e\,(k_p\,k_r + h_p\,h_r + d_p\,d_r)$
K_{i1}: invoiced # kilometers truck type 1 ($K_{i1} - K_{i6}$ for 6 different types of trucks) H_{i1}: invoiced # hours truck type 1 ($H_{i1} - H_{i6}$ for 6 different types of trucks) D_i: invoiced # drivers' hours k_{r1}: rate per kilometer for truck type 1 ($k_{r1} - k_{r6}$ for 6 different types of trucks) h_{r1}: rate per truck hour for truck type 1 ($h_{r1} - h_{r6}$ for 6 different types of trucks) d_r: rate per driver hour	K_a: standard # kilometers given actual routes H_a: standard # hours given actual routes + standard times for loading and unloading $k_{r,i}$: average invoiced rate per kilometer $h_{r,i}$: average invoiced rate per hour of a truck $d_{r,i}$: invoiced rate per driver hour	P_a: standard # pallets based on actual sales volume k_p: standard # kilometers per pallet h_p: standard # hours of trucks per pallet d_p: standard # drivers' hours per pallet k_r: standard rate per kilometer h_r: standard rate per hour of a truck d_r: standard rate per driver hour *Other symbols used:* $k_{p,a}$: actual # kilometers per pallet $h_{p,a}$: actual # hours of trucks per pallet $d_{p,a}$: actual # drivers' hours per pallet	P_e: standard # pallets based on estimated sales volume

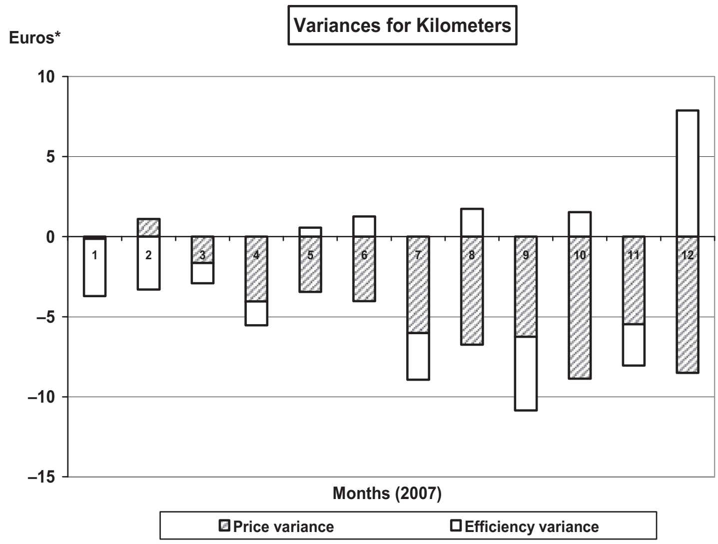


the average invoiced rate per kilometer: $k_r - k_{r,i}$), multiplied by the actual quantity of the activity (e.g., the number of kilometers K_a).

The calculation of the *efficiency variance* was more complicated. Central to the calculation of the budget were three ratios: the standard number of kilometers, hours of trucks, and drivers' hours per pallet. The actual outcomes of these ratios can also be calculated, namely by dividing the actual number of kilometers or hours by the actual number of pallets. The differences between the standard ratios and the actual ratios reflect differences in efficiency, and these were multiplied by the standard rate and the actual number of pallets. (The symbols are defined in Figure 2, e.g., for kilometers the efficiency variance equals $(k_p - k_{p,a})\,k_r\,P_a$, where k_p is the standard number of kilometers per pallet, $k_{p,a}$ is the actual number of kilometers per pallet, k_r is the standard rate per kilometer, and P_a is the standard number of pallets based on the actual sales volume.)

The PMS developed by the company resulted from an experimental process with many explanations, clarifications, prototypes, and iterations. A first version of a variance analysis was presented in a meeting on November 11, 2005, when the controller commented, "If we can have this chart every month, that would already be very nice." It was agreed that "we need to be creative from the beginning as to whether this is the chart. And there needs to be continuous discussion with Bert and Martin [the transportation manager and the distribution manager] about what they would like different." The transportation manager concluded, "If I can get something like this, then I'm happy."

Figure 3 Presentation of the transportation cost variance analysis.



Notes:

*Based on the actual data for 2007, where the values on the *y* axis are divided by a constant in order to conceal exact amounts.

Over time, issues regarding data and calculations were identified and resolved, the graphical representation was evaluated and changed, and discussions were held about how to make the spreadsheet easy to use so that the transportation manager could update it monthly. Detailed descriptions of specific episodes of experimentation with prototypes are provided in the following sections, and these are illustrative of the granularity with which organizational members from different departments were learning from each other. These examples also show the level of detail at which the researchers collected information; they were formed by combining data from different sources (documents, information systems, interviews, meetings, as indicated in Table 1) and working with company employees over an extended period of time.

The evidence presented earlier will demonstrate how user participation and knowledge exchange benefited from experimentation with prototypes, because this process involved (1) representing, (2) learning, and (3) transforming knowledge. The observations will also show how knowledge integration was reinforced because experimentation was conducted using *actual data*, which above we described as experimentation with contextualized data. The experimental prototypes formed the basis for revising the information and resolving discussions, as the various managers and professionals knew they *needed to agree*, which resulted in joint ownership of experimentation. From the beginning it was clear that the *user* (the transportation manager, in this case) — not an accountant — would manage the system once it was in place, which above was described as user reporting.

Representing knowledge: Actual costs in the general ledger

Constructing the new accounting information required all participants to understand actual costs as recorded in the general ledger. As this knowledge was shared, it also became represented (embedded) in the prototype of the new accounting information (see the column *general ledger* in Figure 2). The transportation firm invoiced *standard* amounts of kilometers and hours. The standard amounts of kilometers and hours of trucks for the *actual* routes driven were determined by the routing software system. Standard times for loading and unloading trailers were also included in that system. There were six different types of trucks and consequently six different amounts of kilometers ($K_{i1} - K_{i6}$), hours of trucks ($H_{i1} - H_{i2}$), rates per kilometer ($k_{r1} - k_{r6}$), and rates per hour ($h_{r1} - h_{r6}$). There was one amount for the number of drivers' hours (D_i) and one hourly rate for drivers (d_r). The rates were set for an entire year, with one exception because the rate per kilometer was adjusted monthly for the actual price of diesel fuel. The process of representing this knowledge in the new system involved clarifying several issues that came up in the discussions around the prototype.

Example One: Truck hours (H_i) and driver hours (D_i).

Hours of trucks and hours of driver were the same physically, but not in terms of invoicing. This difference became apparent when data for a period of 12 months in the routing software indicated 3,683,638 minutes of transportation time and 3,453,411 minutes for driver time. This observation sparked a discussion as to what the difference represented. Drivers were legally required to rest, so it was most efficient when a break took place while a truck was being unloaded or loaded at a customer's site, so that resting would not require additional truck hours. When this was discussed, the following question came up: Is the time spent resting invoiced, or instead included as a surcharge in the rate? This question led to addressing some detailed questions about invoicing and budgeting, and it became clear that resting times should not be invoiced; it was verified that these hours were not invoiced. The following question was also raised: If the actual time spent resting is higher than normal, is that a cost for the company or for the transportation firm? It became clear that a standard difference between hours for trucks and drivers of 6.25 percent was used for invoicing to the company, and positive or negative variances of resting times were the transportation firm's risk.

The information described in this example was only shared and integrated because different people were involved in the process of developing the new accounting information. For that purpose, information from different domains was collected and used in trying to construct a work-in-progress version of the variance analysis system. This caused discovery of an unreconciled difference between the numbers of hours of trucks and drivers, which led to knowledge integration: people explaining more to each other about the contract, invoicing, and operational processes. The combined knowledge was then used in the construction of the new accounting information.

Example Two: Selling empty crates.

In one of the project team meetings (January 18, 2006) an issue arose as to what happens if empty crates are sold to a customer. It became clear that the determination of the budget was not well understood. "What happens if we sell a pallet with empty crates?" the distribution manager asked. "Does that ever happen?" one of the researchers asked. The manager replied, "Yes, sometimes customers need empty crates in their warehouse or in their stores for handling single bottles. So how does this work in our budget? Is it counted as an additional pallet sold and thus included in the cost budget? Or are 'pallets sold' only about selling full bottles?" After further discussion by the group, the controller concluded that he needed to investigate the issue and would explain it at a later meeting.

This is an extremely detailed issue, and its relevance for this research is not in the precise meaning of it, but in the fact that it was raised at all. It arose in a dialogue between controllers and operations managers — not in a meeting among finance professionals such as controllers and the auditors. This example again illustrates how knowledge became shared as a result of different people being involved in the process of developing the new accounting information.

Example Three: Loading times.

The standard time in the controller's budgeting system for loading a trailer at the manufacturing site was 45 minutes. However, during one of the meetings (February 16, 2006) it became apparent that the managers from Transportation and Distribution had used a loading time of 25 minutes in the routing software system. Again, the discovery of a difference in key data triggered an exchange of knowledge, in this case after the meeting. When the researcher and employees from the controller's office examined the data in the routing software system, they noticed that the column for loading and unloading times contained 25 minutes and not 45 minutes. They checked data from earlier years in the same system, where it was recorded as 45 minutes. It became clear that loading times had been adjusted to 25 minutes in the agreements with the transportation firm, and this was the amount used for invoicing purposes. However, the amount had not been adjusted in the budgeting system. This discrepancy became apparent through the process of experimentation. The budgeting system was eventually corrected. This example illustrates a range of processes, from simply representing knowledge to discovering differences and subsequently transforming knowledge.

Representing knowledge: Actual costs

The *actual costs* column indicated in Figure 2 is information that the company had not produced before this study. The actual cost that should be invoiced by the transportation firm could be determined by using data on all routes driven and standard amounts of kilometers and hours for these routes. These actual costs as calculated and the invoiced costs in the general ledger (the "real" actual costs) were not clearly distinguished early in the development process. It was the process of experimentation that led to a conceptual understanding of these different kinds of costs.

Example Four: Different kinds of actual costs.

Initially there were small variances between these different types of costs. This occurred, for example, because in January 2006 the costs of one truck that was still owned by the company were in the general ledger account for actual transportation costs, but these internal costs should be in another account. The same occurred for the costs of two drivers that were still on the company's payroll (the outsourcing of transportation had been implemented gradually). Further, some delayed invoices from the transportation firm were booked in a different month than when the transportation had actually been carried out. After some adjustments to the procedures, these variances were eliminated. It was the representation of knowledge while experimenting with prototypes of the new accounting information that led to a conceptual understanding of these different kinds of costs. When this distinction became clearer, the controller advocated incorporating both kinds of cost numbers into the system because that made it possible to verify the invoiced amounts. Thus, the exchanged knowledge became embedded in the accounting prototype.

Example Five: Six different rates.

The costs recorded in the general ledger were based on six different rates for truck kilometers and hours. For the calculation of the actual cost, the average rate was used (total cost of trucks ÷ total kilometers). Therefore, in principle, a price difference would be possible if the mix of trucks used differed from the mix assumed in the budget. However, the mix of trucks used rarely changed, and this variance amount was negligible. This is another example of information that was made explicit and shared during — and because of — the development process.

Representing, learning, and transforming knowledge: The budget

This element of the PMS is shown in the budget column in Figure 2. The starting point for the budget was an estimation of the next year's sales volumes for each product (dozens of different products were distinguished). These sales volumes were translated into the number of standard pallets to be transported (P_e), based on the size of each product.[3] Furthermore, transportation to destinations other than customers' locations was also translated into the equivalent number of standard pallets. For example, some products first went to a repackaging firm for special promotional packaging, which was located closer to the manufacturing site than the average customer, and 100 pallets transported to the repackaging firm was translated into an equivalent number of 23 standard pallets. So, the number of pallets to be transported (P_e) was considered the main cost driver in the budgeting system.

Based on the total expected number of pallets, the number of kilometers and hours were estimated using three operational ratios: the standard number of kilometers per pallet (k_p); the standard number of truck hours per pallet (h_p); and the standard number of drivers' hours per pallet (d_p). The estimated numbers of kilometers and hours were multiplied by the agreed-upon rates (not differentiated for different types of trucks). These operational ratios were based on the actual outcomes of the previous year, so it was assumed that standard operational characteristics would apply in the next year, such as the mix of different types of trucks, the utilization of trucks, and the distance to customers.

Experimentation with this part of the PMS involved not only representing knowledge, but it also led to the discovery of differences, the resolution of which resulted in learning and transforming knowledge.

Example Six: Operational parameters in the budget.

The controller's office downloaded data from the routing software regarding the number of hours, kilometers, and pallets transported; reworked the data; and calculated the ratios for the new budget (kilometers and hours per pallet: k_p, h_p, and d_p). While experimenting with the new PMS, the employees in the controller's office gained a better understanding of the data they downloaded; as a result they identified and subsequently corrected some inaccuracies in the budgeting process.

Example Seven: Conversion factors.

The controller's office used data from the route planning system to calculate the conversion factors for translating transportation to destinations other than customers into the equivalent number of standard pallets. When discussing the budget, the transportation manager looked at these factors in the budgeting system in more detail, and he observed that in some cases a factor of 0.74 was used. He believed that this was too high and that something in the system had to be incorrect. This led to further discussions about the data in the routing system, and the problem was corrected.

Transforming knowledge: The flexible budget

The flexible budget (Figure 2) was based on the same standards as those underlying the original budget, but was adjusted for the actual sales volume and the actual number of pallets. The actual number of transported pallets (P_a) was multiplied by the same three ratios as in the original budget (standard number of kilometers k_p, hours of trucks h_p, and driver's hours d_p per pallet) to determine the allowable amounts for kilometers and hours.

3. For example, for one particular product, one crate contains 24 0.33 liter bottles, and one pallet contains 70 crates; one pallet thus equals a sales volume of 554 liters.

Multiplying by the three standard rates (k_r, h_r, d_r) yields the flexible budget. The actual number of pallets was a central concept for developing the new accounting information, but employees in the controller's office and in the transportation department understood this differently. The next example will clearly illustrate the implications of effective experimentation proposed in section 3 and show the level of granularity and specificity at which knowledge integration took place.

Example Eight: Actual number of pallets.

At first, it seemed that the route planning system, which planned transportation routes on the basis of customer orders, contained the actual number of pallets transported. Certain data in the system were actually labeled as such. This was verified with the transportation manager, who had the most knowledge about the route planning system, and he explained that these data referred to the number of *physical* pallets that were on each truck when it left the manufacturing site. He doubted that this number was the actual number of pallets transported, as in the budgeting system, but he was not sure about the connection between the two kinds of data.

A key to this experimental process was the use of *actual data* taken from *actual information* systems as input for experimentation to stimulate representing and learning. Early prototypes of the new performance measures were constructed with historic data for the month of August 2005; the actual number of pallets according to the routing system was 44,586 for that month while the number of pallets according to the budgeting system was 25,212. Follow-up on this inconsistency with employees from transportation, sales, and the controller's office (which required several rounds of probing) revealed that these numbers represented different concepts of "actual pallets".

Discussions with the controller's office made it clear that in the budgeting system "pallets" was a more abstract concept. It was used as the *cost driver* to budget the cost of transportation and this was a measure of transportation activity level. Above we described that the expected sales volume was translated into the expected number of pallets to be transported, and for the flexible budget the actual sales volume was translated into the "actual" number of pallets in the same way. Also, the actual transportation to destinations other than customers' locations was translated into the equivalent number of standard pallets. So, despite the terminology, "actual number of pallets" in the budgeting system had nothing to do with the transported number of physical pallets recorded in the route planning system. Neither the controller's office nor the transportation department was initially aware of these differences in the meaning of the term "pallets". The differences only became apparent and were resolved because the process of experimentation with prototypes of the new accounting information was based on actual data.

Example Nine: Flexible budget calculation.

Around December 2005, using data from August 2005, the prototype of the variance analysis tool retrospectively produced an amount for the flexible budget of August 2005 that differed significantly from the flexible budget of August 2005 calculated by the controller's department. This was subsequently investigated, and it turned out that a mistake had been made in the flexible budgeting system, caused by using the wrong data concerning the actual number of pallets. In this example, the controller's department and the researchers were involved (rather than employees from other departments), but again knowledge representation, learning, and transformation occurred because of the prototyping exercise using actual data and information systems.

As a summary, in this section on designing and implementing the new accounting information, the examples show in detail that experimentation with prototypes can enhance user participation and knowledge exchange, with stronger effects when those prototypes are based on actual data and information systems. In section 3, this was called *experimentation with contextualized data*. In several examples, using actual data led to the observation that "something must be wrong here", which was typically the starting point for further information exchange. We suggest that experimentation with contextualized data reinforces the antecedents of knowledge integration identified by Carlile 2002, and especially enhances *learning about differences and interdependencies*. This relation is indicated in Figure 1.

Furthermore, throughout the process of experimentation with the new accounting information, it was clear that the different parties involved needed to agree on the final design of the PMS. Because the information was intended for the transportation manager, and not for reporting his performance to higher management levels, he was entitled to make decisions on the PMS. Also, the controller's office had the responsibility to improve the accuracy of budgets and the analysis of actual costs provided to higher-level managers. Consequently, while this new PMS was not for higher-level managers, the controller's office was entitled to make decisions on the accuracy and consistency of it. The operational and accounting employees involved in the experimentation understood and respected these responsibilities, and it was quite clear that they needed to reach agreement on the PMS design. The necessity to reach agreement added to the intensity with which representing, learning, and transforming took place. In section 3, this characteristic was summarized as *joint ownership of experimentation*. We expect this characteristic to reinforce knowledge integration generally and more strongly through the *transformation of knowledge* because of the urgent need to reach agreement. This relation is indicated in Figure 1.

Regularly reporting and using the new accounting information

At the end of the study, the company started using the PMS they had designed. The system was documented and the transportation manager started producing the accounting numbers. Every month, the actual costs and the flexible budget were determined, and the variance was split into flexible budget, price, and efficiency variances. Results were also presented graphically.

From the start, it was the explicit intention that the transportation manager would produce and report the new accounting information. In section 3 we described this as *user reporting*. When the first version of the variance analysis was presented, it was agreed that a requirement for any new information to be developed was that it should be producible in a reasonable amount of time, taking a maximum of two hours of the transportation manager's time. The prospect of user reporting stimulated in-depth participation of the transportation manager. He was aware that he needed to understand the PMS in more detail, compared to when the controller's office produced the information. He also realized that he was going to spend time working with this PMS on an ongoing basis, which gave him more incentive to participate and make the information worthwhile and relevant to his needs.

After this study, the company reflected on the experiences during the first year of using the new accounting information. The system was still in use and had helped in the monitoring and control of transportation cost. Compared to a situation in which the company would only contract on a fixed cost per pallet, the company better understood the factors that determined actual costs. When the negotiations for renewal of the contract took place (2007), the system enabled the company, together with the transportation firm, to identify efficiency improvements (e.g., changing processes so that trucks could be used more hours every day) which led to significant cost reductions. Some limitations of the system were also experienced. In 2007 the implementation of a packaging design change caused inefficiency, because it temporarily made the goods flow to customers and the return flow of packaging materials more unbalanced. However, the system could not isolate this specific cause of inefficiency. Furthermore, the controller's office felt that the system helped improve its daily cooperation with the transportation department, but sometimes they were so involved in the operations, that their management control role became less distinct.

6. Discussion and conclusion

This study describes a process for developing an enabling PMS. Constructing this accounting information required detailed knowledge integration, involving operations managers and employees in the controller's office. Why did this happen successfully? Knowledge about transportation activities, operational data, and costs were required from different functions and from different information systems, such as the routing system, the ERP system, and the budgeting system. The required level of detail and specificity of knowledge was extensive and spread across different employees. Consistent with the previous literature on the adoption of accounting innovations (McGowan and Klammer 1997; Abernethy and Bouwens 2005), knowledge integration and cooperation between accountants and users (Chenhall and Langfield-Smith 1998; Emsley 2005) were important in this setting, but the research question we addressed was *how* this could be facilitated.

Knowledge integration was stimulated through experimentation. The prototypes consisted of numerous versions of spreadsheets "under construction". Also the researchers produced presentations and documents about the prototypes, to explain and guide discussions. The data used in these prototype versions were taken from the actual databases and information systems that were expected to become routinely used. The calculations in a particular version represented the conceptualization of the variance analysis at that point in the process. Graphs presenting the results were included in the early versions to create discussion about the structure and appearance of the output. The final version of this spreadsheet model was the PMS that became routinely used. The creation and discussion of prototypes of the PMS led to asking many questions across functional boundaries, because comparisons between data from different sources, and comparisons between existing data and the outputs generated by a prototype, often revealed inconsistencies. The questions were related to operational processes, procedures for recording operational events in information systems, the meaning of data in those systems, details of the contract for transportation activities, invoicing conditions, and so on. These were "how-does-it-work" questions that individuals had not asked before, although they had been working at the company in their current role for several years. Asking questions led to explaining, recognizing gaps in understanding, and investigating queries. Ideas about the new accounting information represented in prototypes prompted questions about the proposed information, responding to each other's suggestions, coming up with new ideas, listing issues that needed to be resolved, and further developing agreed-upon concepts.

This process of experimentation was made even more intense because from very early on the experimentation was conducted with actual data taken from actual information systems, which highlighted key differences in understanding that needed be understood and resolved. Furthermore, the experimentation resulted in effective knowledge exchange because it was clear from the start that employees from the controller's office and the transportation department needed to agree on the PMS design. It was also explicit that the transportation manager was going to work with the new PMS, not only as a user but also as a future preparer of the information. Previous studies have focused on the difficulties of selecting and interpreting appropriate performance metrics in more complex settings (Cavalluzzo and Ittner 2004; Chenhall and Morris 1995). This study demonstrates how, even in settings that are relatively straightforward in terms of modeling operational processes and measuring outcomes, the construction of enabling accounting information at an operational level involves detailed exchange and integration of knowledge.

The contribution of this study is to provide a more nuanced understanding of *how* experimentation with prototypes can be effective for the development of enabling PMS. Knowledge integration (D'Adderio 2001) and experimentation (Thomke et al. 1998) are important for innovation, and user participation is important for the adoption of accounting innovations (McGowan and Klammer 1997; Abernethy and Bouwens 2005). But the accounting literature has little to say about facilitating the process of involving users, exchanging knowledge, integrating it in new accounting information, and experimenting to make it enabling accounting information.

This study documents several ways in which prototypes can be used to induce experimentation and knowledge integration in the development process of enabling accounting information. The first is representing knowledge, in that experimentation requires participants to represent their knowledge in terms of a prototype of the PMS under development. A prototype constitutes an "object" used to generate and answer questions, with the results becoming embedded into the next version of the prototype system. Then learning and transforming knowledge can occur, because by explicating knowledge in the form of prototypes, individuals identify and learn about their differences and interdependencies. The identification and resolution of differences in understanding may happen in several rounds of experimentation, whereby transformed knowledge becomes incorporated in a new version of an experimental prototype, which may lead to new questions and discussions. Furthermore, these characteristics are reinforced by experimentation with contextualized data. The use of real data from actual information systems increases the depth and specificity with which users need to represent their knowledge and learn from each other. In particular, these data are likely to lead to the discovery of more differences and interdependencies, and this is also represented in Figure 1.

The representation, learning, and transformation of knowledge also becomes more compelling when there is joint ownership of experimentation and user reporting. Joint ownership indicates that both accountants and nonaccountants need to reach a common understanding and agreement on the conceptualization and actual implementation of the PMS. This requires the different functions involved to express their knowledge in more detail, to consider more carefully the knowledge brought forward by others, and to put more energy into convincing, listening, and looking for a solution. That is why we expect that joint ownership reinforces the characteristics of effective knowledge integration. In particular, joint ownership is likely to lead to enhanced knowledge transformation, and this is also represented in Figure 1. User reporting means that a nonaccounting user will prepare performance measurement information. As a result, the user needs to better understand the new information and is likely to put in more effort to influence outcomes so that the information is relevant and justifies the time required for preparation.

A second (albeit smaller) contribution of this study is to document an actual PMS for managing transportation costs, based on variance analysis. This has received little attention in the literature, with Gaffney, Gladkikh, and Webb 2007 as a notable exception. To the extent that variance analysis has been addressed in the literature, the emphasis has been on manufacturing settings. For example, Emsley (2001) investigates how variance analysis can improve problem solving, using a longitudinal case study in a manufacturing company; Balakrishnan and Sprinkle (2002) discuss an approach for teaching variance analysis in connection with capacity costing, using a manufacturing setting; textbooks by Horngren, Bhimani, Datar, and Foster 2005 and Drury 2004 provide teaching examples in manufacturing. The application of variance analysis to other kinds of activities is relatively unexplored (Gaffney et al. 2007). However, there are conceptual differences between manufacturing and transportation, such as the basis for establishing the "standard cost" in transportation, the "unit" for measuring volume (kilometers, pallets, truckloads), and distinct ways to disaggregate the overall variance into separate components. The study by

Gaffney et al. (2007) demonstrates that the conceptualization of variance analysis of distribution costs depends on specific operational characteristics, such as the presence of a network of plants and warehouses, and the possibility of using alternative modes of transportation (trucks and rail) and alternative carriers. Ernst & Whinney (1983) provides another exception, describing an example of variance analysis of transportation costs, which includes a cost variance analysis due to mix changes (i.e., full truckloads versus less-than-full truckloads).

This study has a number of limitations. It is difficult to draw empirical generalizations from a single case study, even though it is based on hundreds of observations. We paid close attention to comparing observations in this particular case with theoretical notions about experimentation, user participation, and knowledge integration, thus striving for theoretical generalization. We suggest that the following characteristics, which formed the theoretical starting point for the current study, may be particularly relevant for theoretical generalization: a focus on operations close to the shop floor, an enabling intent for new accounting information, room for experimentation, and the need for user participation and cross-functional knowledge integration. We also acknowledge that action research may have limitations in terms of possible biases (e.g., selective perception and interpretation). However, we are confident that our account accurately describes the development process and the resulting PMS and that the study's validity is high regarding the analysis of prototypes as integrating devices. Not only were we able to collect diverse types of qualitative data (e.g., notes from interviews and meetings, data from information systems, documents, and emails), but more importantly, being involved as action researchers provided several strengths (Jönsson and Lukka 2007). First, by being actively involved in the development of the PMS, we obtained firsthand information on the process, including: who participated, how people reacted to prototypes, which questions were raised, what information was exchanged, what changes to a particular version of a prototype were proposed, and which data could be found in which information system. Second, we achieved strong involvement from organizational members. They expected results that were of practical relevance, and this expectation provided an incentive to be involved and spend time with the researchers. Participants were engaged, challenged ideas, and provided feedback on results. Third, this type of action research allowed an empirical test of ideas implemented at an actual organization. We could answer questions such as: Was the developed system for managing the cost of transportation considered useful? Was a process of experimentation effective? Did the PMS remain in use?

A suggestion for future research is to more fully develop the understanding of experimentation by investigating the approach in other empirical contexts and to pose additional questions, such as: What are the antecedents of a development process of experimentation with prototypes? What are the consequences, in terms of performance? Researching such questions may require gathering more standardized, less in-depth data (qualitative and quantitative). Future research could also look for connections with positive research themes. Experimentation as it is developed in this study is about utilizing knowledge dispersed across different employees and functional areas in order to create a more valid PMS. Doing so is not trivial, even in relatively simple operational settings (Ahrens and Chapman 2004; Lillis 2002). We sought to better understand how PMS development can be conducted effectively from the perspective of enabling PMS. However, we can look at the creation of a more valid PMS from an agency perspective: better quality performance measures can be seen as reducing measurement error, which is related to organizational design choices (Abernethy, Bouwens, and van Lent 2004; Moers 2006). Moers (2006) investigates the relationship between the quality of financial performance measures and the extent of delegation of decision rights. Delegation provides the agent with more degrees of freedom to make tradeoffs, making it more appropriate to use aggregate financial measures rather than specific nonfinancial measures. However, this requires the aggregate measures to have good measurement properties. Moers finds that if financial performance measures have good measurement properties, then using these measures for incentive purposes increases delegation. We did not frame our research in this light, but experimentation might also be an antecedent to reduced measurement error and organizational design choices. Experimentation with a performance measure, to the extent it reduces measurement error, may also be associated with more delegation of the trade-offs that are encompassed by that measure. This may also apply to detailed activities lower in the organization. For example, a more valid measure of on-time delivery may be associated with more freedom for managers to execute activities to achieve a particular performance objective for delivery activities. Future research into development processes of PMS could investigate such connections between the enabling and controlling objectives of accounting systems.

To conclude, while user participation and cross-functional knowledge integration involving accountants and operations managers is difficult, it is vital for the development of enabling accounting information. This study provides further insights into how such processes may be conducted effectively. Developing enabling accounting information is a multidisciplinary endeavor, and experimentation with prototypes may enable knowledge integration for developing enabling accounting information.

References

Abernethy, M. A., and J. Bouwens. 2005. Determinants of accounting innovation implementation. *Abacus* 41 (3): 217–40.

Abernethy, M. A., J. Bouwens, and L. van Lent. 2004. Determinants of control system design in divisionalized firms. *The Accounting Review* 79 (3): 545–70.

Adler, P. S., and B. Borys. 1996. Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly* 41 (1): 61–89.

Ahrens, T. A., and C. S. Chapman. 2002. The structuration of legitimate performance measures and management: Day-to-day contests of accountability in a U.K. restaurant chain. *Management Accounting Research* 13: 151–71.

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

Ahrens, T. A., and C. S. Chapman. 2006. Doing qualitative field research in management accounting: Positioning data to contribute to theory. *Accounting, Organizations and Society* 31 (8): 819–41.

Anderson, S. W. 1995. A framework for assessing cost management system changes: The case of activity-based costing implementation at General Motors, 1986–1993. *Journal of Management Accounting Research* 7: 1–65.

Anderson, S. W., and K. Sedatole. 1998. Designing quality into products: The use of accounting data in new product development. *Accounting Horizons* 12 (3): 213–33.

Anderson, S. W., and S. M. Young. 1999. The impact of contextual and process factors on the evaluation of activity-based costing systems. *Accounting, Organizations and Society* 24 (7): 525–59.

Atkinson, A. A., and W. Shaffir. 1998. Standards for field research in management accounting. *Journal of Management Accounting Research* 10: 41–68.

Balakrishnan, R., and G. B. Sprinkle. 2002. Integrating profit variance analysis and capacity costing to provide better managerial information. *Issues in Accounting Education* 17 (2): 149–61.

Bechky, B. A. 2003. Sharing meaning across occupational communities: The transformation of understanding on the production floor. *Organization Science* 14 (3): 312–30.

Carlile, P. R. 2002. A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science* 13 (4): 442–55.

Carlile, P. R. 2004. Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization Science* 15 (5): 555–68.

Cavalluzzo, K. S., and C. D. Ittner. 2004. Implementing performance measurement innovations: Evidence from government. *Accounting, Organizations and Society* 29 (3–4): 243–67.

Chapman, C. S. 1998. Accountants in organisational networks. *Accounting, Organizations and Society* 23 (8): 737–66.

Chapman, C. S., and L.-A. Kihn. 2009. Information system integration, enabling control and performance. *Accounting, Organizations and Society* 34 (2): 151–69.

Chenhall, R. 2008. Accounting for the horizontal organization: A review essay. *Accounting, Organizations and Society* 33 (4–5): 517–50.

Chenhall, R., and K. J. Euske. 2007. The role of management control systems in planned organizational change: An analysis of two organizations. *Accounting, Organizations and Society* 32 (7–8): 601–37.

Chenhall, R., and K. Langfield-Smith. 1998. Factors influencing the role of management accounting in the development of performance measures within organizational change programs. *Management Accounting Research* 9 (4): 361–86.

Chenhall, R., and D. Morris. 1995. Organic decision and communication processes and management accounting systems in entrepreneurial and conservative business organizations. *Omega* 23 (5): 485–97.

Coghlan, D. 2001. Insider action research projects: Implications for practising managers. *Management Learning* 32 (1): 49–60.

Coughlan, P., and D. Coghlan. 2002. Action research for operations management. *International Journal of Operations and Production Management* 22 (2): 220–40.

D'Adderio, L. 2001. Crafting the virtual prototype: How firms integrate knowledge and capabilities across organisational boundaries. *Research Policy* 30 (9): 1409–24.

Davila, A., and D. Oyon. 2008. Cross-paradigm collaboration and the advancement of management accounting knowledge. *Critical Perspectives on Accounting* 19 (6): 887–93.

Davis, F. D., and V. Venkatesh. 2004. Toward preprototype user acceptance testing of new information systems: Implications for software project management. *IEEE Transactions on Engineering Management* 51 (1): 31–46.

Dent, J. F. 1991. Accounting and organizational cultures: A field study of the emergence of a new organizational reality. *Accounting, Organizations and Society* 16 (8): 705–32.

Derfuss, K. 2009. The relationship of budgetary participation and reliance on accounting performance measures with individual-level consequent variables: A meta-analysis. *European Accounting Review* 18 (2): 203–39.

Drury, C. 2004. *Management and cost accounting*, 6th ed. London: Thomson Learning.

Dubois, A., and L.-E. Gadde. 2002. Systemic combining: An abductive approach to case research. *Journal of Business Research* 55 (7): 553–60.

Eisenhardt, K. M., and J. A. Martin. 2000. Dynamic capabilities: What are they? *Strategic Management Journal* 21 (10–11): 1105–21.

Emsley, D. 2001. Redesigning variance analysis for problem solving. *Management Accounting Research* 12: 21–40.

Emsley, D. 2005. Restructuring the management accounting function: A note on the effect of role involvement in innovativeness. *Management Accounting Research* 16: 157–77.

Eraut, M. 2007. Learning from other people in the workplace. *Oxford Review of Education* 33 (4): 403–22.

Ernst & Whinney. 1983. *Transportation accounting and control: Guidelines for distribution and financial management*. New York: The National Association of Accountants, and Oak Brook, IL: The National Council of Physical Distribution Management.

Eva, M. 2001. Requirements acquisition for rapid applications development. *Information & Management* 39 (2): 101–07.

Ezzamel, M., S. Lilley, and H. Willmott. 2004. Accounting representation and the road to commercial salvation. *Accounting, Organizations and Society* 29 (8): 783–813.

Free, C. 2007. Supply-chain accounting perspectives in the UK retail sector: Enabling or coercive collaboration? *Contemporary Accounting Research* 24 (3): 897–933.

Frow, N., D. Marginson, and S. Ogden. 2005. Encouraging strategic behaviour while maintaining management control: Multi-functional project teams, budgets, and the negotiation of shared accountabilities in contemporary enterprises. *Management Accounting Research* 16: 269–92.

Fry, T., D. C. Steele, and B. A. Saladin. 1995. The role of management accounting in the development of a manufacturing strategy. *International Journal of Operations and Production Management* 15 (12): 21–31.

Gaffney, K., V. Gladkikh, and R. A. Webb. 2007. A case study of a variance analysis framework for managing distribution costs. *Accounting Perspectives* 6 (2): 167–90.

Galbraith, J. 1973. *Designing complex organizations*. Reading, MA: Addison-Wesley.

Gordon, V. S., and J. M. Bieman. 1994. Rapid prototyping: Lessons learned. *IEEE Software* 12 (1): 85–95.

Gosselin, M. 2007. A review of activity-based costing: Technique, implementation, and consequences. In *Handbook of management accounting research*, ed. C. S. Chapman, A. G. Hopwood, and M. D. Shields, 641–71. Oxford, UK: Elsevier.

Hedberg, B., and S. Jönsson. 1978. Designing semi-confusing information systems for organizations in changing environments. *Accounting, Organizations and Society* 3 (1): 47–64.

Horngren, C. T., A. Bhimani, S. M. Datar, and G. Foster. 2005. *Management and cost accounting*, 3rd ed. Harlow, UK: Pearson Education.

Hunton, J. E., and J. D. Beeler. 1997. Effect of user participation in systems development: A longitudinal field experiment. *MIS Quarterly* 21 (4): 359–88.

Ittner, C. D., and D. F. Larcker. 2003. Coming up short on nonfinancial performance measurement. *Harvard Business Review* 81 (11): 88–95.

Jönsson, S., and A. Grönlund. 1988. Life with a sub-contractor: New technology and management accounting. *Accounting, Organizations and Society* 13 (5): 512–32.

Jönsson, S., and K. Lukka. 2007. There and back again: Doing interventionist research in management accounting. In *Handbook of management accounting research*, eds. C. S. Chapman, A. G. Hopwood, and M. D. Shields, 373–97. Oxford, UK: Elsevier.

Jørgensen, B., and M. Messner. 2010. Accounting and strategising: A case study from new product development. *Accounting, Organizations and Society* 35 (2): 182–204.

Kaplan, R. S. 1998. Innovation action research: Creating new management theory and practice. *Journal of Management Accounting Research* 10: 89–118.

Kaplan, R. S., and D. P. Norton. 2006. How to implement a new strategy without disrupting your organization. *Harvard Business Review* 84 (3): 100–09.

Kasanen, E., K. Lukka, and A. Siitonen. 1993. The constructive approach in management accounting research. *Journal of Management Accounting Research* 5: 243–64.

Krumwiede, K. R. 1998. The implementation stages of activity-based costing and the impact of contextual and organizational factors. *Journal of Management Accounting Research* 10: 239–77.

Labro, E., and T.-S. Tuomela. 2003. On bringing more action into management accounting research: Process considerations based on two constructive case studies. *European Accounting Review* 12 (3): 409–42.

Leenders, M. A. A. M., and B. Wierenga. 2002. The effectiveness of different mechanisms for integrating marketing and R&D. *Journal of Product Innovation Management* 19 (4): 305–17.

Levina, N., and E. Vaast. 2005. The emergence of boundary spanning competence in practice: Implications for implementation and use of information systems. *MIS Quarterly* 29 (2): 335–63.

Lillis, A. M. 2002. Managing multiple dimensions of manufacturing performance: An exploratory study. *Accounting, Organizations and Society* 27 (6): 497–529.

Malina, M. A., H. S. O. Nørreklit, and F. H. Selto. 2007. Relations among measures, climate of control, and performance measurement models. *Contemporary Accounting Research* 24 (3): 935–82.

McDonough, E. F. III. 2000. Investigation of factors contributing to the success of cross-functional teams. *Journal of Product Innovation Management* 17 (3): 221–35.

McGowan, A. S., and T. P. Klammer. 1997. Satisfaction with activity-based cost management implementation. *Journal of Management Accounting Research* 9: 217–37.

McKinnon, S. M., and W. J. Bruns, Jr. 1992. *The information mosaic*. Boston: Harvard Business School Press.

Mitchell, V. L. 2006. Knowledge integration and information technology project performance. *MIS Quarterly* 30 (4): 919–39.

Moers, F. 2006. Performance measure properties and delegation. *The Accounting Review* 81 (4): 897–924.

Nambisan, S., and D. Wilemon. 2000. Software development and new product development: Potentials for cross-domain knowledge sharing. *IEEE Transactions on Engineering Management* 47 (2): 211–20.

Naranjo-Gil, D., and F. Hartmann. 2006. How top management teams use management accounting systems to implement strategy. *Journal of Management Accounting Research* 18: 21–53.

O'Leary, D. E. 1988. Software engineering and research issues in accounting information systems. *Journal of Information Systems* 2 (2): 24–38.

Otley, D. T. 1999. Performance management: A framework for management control systems research. *Management Accounting Research* 10: 363–82.

Otley, D. T., and A. J. Berry. 1994. Case study research in management accounting and control. *Management Accounting Research* 5: 45–65.

Patnayakuni, R., A. Rai, and A. Tiwana. 2007. Systems development process improvement: A knowledge integration perspective. *IEEE Transactions on Engineering Management* 54 (2): 286–300.

Roller, D., O. Eck, and S. Dalakakis. 2004. Knowledge-based support of rapid product development. *Journal of Engineering Design* 15 (4): 367–88.

Rowe, C. 2004. The effect of accounting report structure and team structure on performance in cross-functional teams. *The Accounting Review* 79 (4): 1153–80.

Shields, M. D. 1995. An empirical analysis of firms' implementation experiences with activity-based costing. *Journal of Management Accounting Research* 7: 148–66.

Sillince, J. A. A., and G. M. H. Sykes. 1995. The role of accountants in improving manufacturing technology. *Management Accounting Research* 6: 103–24.

Simons, R. 1991. Strategic orientation and top management attention to control systems. *Strategic Management Journal* 12 (1): 49–62.

Simons, R. 1995. Control in an age of empowerment. *Harvard Business Review* 73 (2): 80–88.

Song, X. M., R. J. Thieme, and J. Xie. 1998. The impact of cross-functional joint involvement across product development stages: An exploratory study. *Journal of Product Innovation Management* 15 (4): 289–303.

Souder, W. E., J. D. Sherman, and R. Davies-Cooper. 1998. Environmental uncertainty, organizational integration, and new product development effectiveness: A test of contingency theory. *Journal of Product Innovation Management* 15 (6): 520–33.

Thomke, S. 1998. Managing experimentation in the design of new products. *Management Science* 44 (6): 743–62.

Thomke, S., E. von Hippel, and R. Franke. 1998. Modes of experimentation: An innovation process — and competitive — variable. *Research Policy* 27 (3): 315–32.

Tidd, J., and K. Bodley. 2002. The influence of project novelty on the new product development process. *R&D Management* 32 (2): 127–38.

Townley, B., D. J. Cooper, and L. Oakes. 2003. Performance measures and the rationalization of organizations. *Organization Studies* 24 (7): 1045–71.

Tuomela, T.-S. 2005. The interplay of different levers of control: A case study of introducing a new performance measurement system. *Management Accounting Research* 16: 293–320.

Vaivio, J. 2004. Mobilizing local knowledge with ‘‘provocative’’ non-financial measures. *European Accounting Review* 13 (1): 39–71.

Vaivio, J. 2006. The accounting of ‘‘The Meeting’’: Examining calculability within a ‘‘fluid’’ local space. *Accounting, Organizations and Society* 31 (8): 735–62.

Van der Veeken, H., and M. Wouters. 2002. Use of accounting information systems by operations managers in a project company. *Management Accounting Research* 13: 345–70.

Van de Ven, A., and P. E. Johnson. 2006. Knowledge for theory and practice. *Academy of Management Review* 31 (4): 802–21.

West, J., and M. Iansiti. 2003. Experience, experimentation, and the accumulation of knowledge: The evolution of R&D in the semiconductor industry. *Research Policy* 32 (5): 809–25.

Wouters, M., and C. Wilderom. 2008. Developing performance measurement systems as enabling formalization: A longitudinal field study of a logistics department. *Accounting, Organizations and Society* 33 (4–5): 488–515.

Wouters, M., F. Wynstra, and J. C. Anderson. 2005. The adoption of total cost of ownership for sourcing decisions: A structural equations analysis. *Accounting, Organizations and Society* 30 (2): 167–91.