Exploring the performance effects of performance measurement system use in maintenance process

Jaakko Tätilä, Pekka Helkiö and Jan Holmström
Department of Industrial Engineering and Management, School of Science, Aalto University, Espoo, Finland

Received 25 July 2013
Revised 23 November 2013
Accepted 19 April 2014

Abstract

Purpose – The intended function of performance measurement is to support the effective management of an organisation and the improvement of organisational performance. However, how performance measurement should be used operationally to support the achievement of improved performance is not self-evident. The purpose of this paper is to examine the operational use of performance measurement in practice and to describe how different use practices contribute to improved performance.

Design/methodology/approach – The authors conducted an exploratory single case study in a maintenance process. Data were collected using a mixed methods approach that encompassed qualitative meetings and interviews (identification of usage practices) followed by a quantitative survey (elaboration of usage practices and their performance effects).

Findings – Three usage practices are relevant: Inspect and Improve, Motivate, and Decision Making. Improved performance is best achieved through motivational and supportive improvement use. Furthermore, performance measurement systems must be designed properly to establish their use.

Research limitations/implications – Being based on a single-case study, the identified usage practices may be limited to field service organisations or other organisations with similar organisational structures. The findings suggest opportunities for further research linking operational performance measurement system use and the body of knowledge on the design and purpose of performance measurement in maintenance processes.

Practical implications – A performance measurement system can be used as a motivational improvement tool in operational level leadership. Upper level management must support its use by designing an understandable and applicable system.

Originality/value – This paper identifies specific usage practices that contribute to improved performance, thereby providing a more detailed view than the usage categories found in the extant literature. The focus is on operational, rather than strategic, level management.

Keywords Performance measurement, Operational performance, Maintenance management, Exploratory case study, Leadership style, Field service

Paper type Case study

1. Introduction

There is more to the successful use of performance measurement systems in maintenance processes than the act of measurement. To be sure, designing an appropriate set of measures and applying them correctly are key to avoiding many known pitfalls (Neely *et al.*, 1997, 2005). However, in the case of misuse or non-use, the intended positive influences of performance measurement on the organisation's performance are lost (Johnston and Clark, 2008; Neely and Bourne, 2000). It is of high practical interest to determine how performance measurement systems should be used operationally to achieve performance effects and what the antecedents of such use are.

The use of performance measurement systems has traditionally been understood as the managerial processes involved in strategic alignment (Johnston and Clark, 2008;

Journal of Quality in Maintenance Engineering
Vol. 20 No. 4, 2014
pp. 377-401

1355-2511
DOI 10.1108/JQME-07-2013-0051

Kaplan and Norton, 1996). More recent research has produced categorisations of day-to-day usage practices (Simons, 2000) and an understanding of the different situations in which performance measurement systems are used (Bourne *et al.*, 2000; Ukko *et al.*, 2007). However, research has not yet identified which usage practices are the most beneficial for performance. The identification of specific practices and their contributions to improved performance would provide a complementary, operational-level view of performance measurement compared to the strategic, higher-level management aspect dominant in the current literature on performance management in maintenance (Parida and Kumar, 2006; Gaiardelli *et al.*, 2007; Muchiri *et al.*, 2011).

The aim of this paper is to explore different practices of performance measurement usage and how they affect performance. Our exploratory case study was guided by the two-part research question:

RQ1. What are the usage practices in field service operations and how are they associated with the performance of the field service unit?

We conducted the case study in the maintenance service unit of a mechanical engineering company. Our mixed methods analysis and data collection began with a qualitative phase to identify the ways in which the supervisors use the performance measurement system. We then expanded and refined the preliminary understanding through an analysis of survey data collected from 149 supervisors working in the case organisation. Based on the analysis we then identify a number of usage practices that generates behavioural effects leading to improved performance of personnel. Furthermore, we found that the design of the performance measurement system, not just how it is used by supervisors, affects behaviours associated with improved performance. Our main contributions are the identification of performance measurement usage practices and the elaboration of the performance effects of the use practices.

2. Literature review

To begin with definitions, we view performance measurement as the quantification of an organisation's goals in terms of efficiency and effectiveness. In simple terms, performance measurement is about quantifying performance (Johnston and Clark, 2008). Following Neely *et al.* (1995) we see performance as having two aspects: efficiency and effectiveness. Efficiency measures how economically the organisation uses its resources and effectiveness captures the extent to which customer requirements are met.

A performance measurement system is a set of measures that are used to measure performance (Neely *et al.*, 1995). A more comprehensive definition includes the processes and procedures that are required to use the system (Franco-Santos *et al.*, 2007; Simons, 2000). Performance measurement system processes consist of activities related to the selection and design of measures, data collection and manipulation, information management, performance evaluation and rewards, and system review (Franco-Santos *et al.*, 2007). However, as Franco-Santos *et al.* (2007) note, the more processes are included in the performance measurement system definition, the more it begins to overlap with other management processes, especially performance management. In this study, we follow Bititci *et al.* (1997) and view the performance measurement system as "the information system which enables the performance management process to function effectively and efficiently" (Bititci *et al.*, 1997, p. 524).

2.1 Constitution of the performance effect of performance measurement systems

Our review of the literature on the topics of performance measurement system use and implementation resulted in a path model of the constitution of their performance effects. Figure 1 illustrates this conceptual model that is based, in particular, on the studies of de Waal (2003), Franco and Bourne (2003), de Leeuw and van den Berg (2011), and Bourne *et al.* (2005). The conceptual model's three propositions link the deployment of a performance measurement system and its operational use with behaviour and performance. The purpose of the conceptual model is to serve as a theoretical foundation for the empirical part of this paper.

The first proposition links the deployment and the operational use of a performance measurement system. The study of de Waal (2003) suggests that critical areas of attention in achieving the use of a performance measurement system include the users' understanding of, attitude towards and responsibility match with a performance measurement system, as well as an organisation's measurement-focused culture, with an emphasis on internal issues. Similarly, Franco and Bourne (2003) identified culture, management leadership, compensation, education and understanding, communication and reporting, reviewing and updating data processes and information technology, measurement framework, environment, and characteristics of measures as the most important factors related to use. Thus, we can focus on the most relevant topics and issues when deploying a performance measurement system, and summarise literature as our first proposition:

P1. The deployment of a performance measurement system has a positive effect on its use.

P2 and *P3* link the use of a performance measurement system with organisational performance. In their study, Bourne *et al.* (2005) found that the interactive use of a performance measurement system was a differentiator between high and average performance. High-performing business units used performance measurement more interactively compared to the simple control approach taken by average-performing business units (Bourne *et al.*, 2005). Interactive use consists of own data collection and analysis, using multiple sources of data, intensive communication, and actions on a wide range of issues (Bourne *et al.*, 2005). Studying a factory floor setting, de Leeuw and van den Berg (2011) show that behavioural effects mediate the positive relationship between performance management practices and performance improvement. Moreover, they argue that the more intensively performance management practices are applied, the greater the performance improvement. Drawing primarily on the study of de Leeuw and van den Berg (2011), we conceptualise a similar mediation relationship in *P2* and *P3*:

P2. The use of a performance measurement system has a positive effect on beneficial behaviours.

P3. Beneficial behaviours have a positive effect on performance.

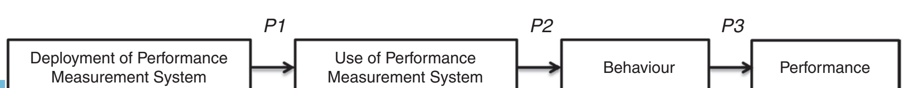


Figure 1. The conceptual model

2.2 Literature on the use of performance measurement systems

Relative to performance measurement system design, the use of performance measurement systems is a much more recent research interest (Franco and Bourne, 2003; Pavlov and Bourne, 2011; de Waal, 2003). In existing research, usage has typically been conceptualised in terms of strategy alignment (Johnston and Clark, 2008; Kaplan and Norton, 1996). Kaplan and Norton (1996) argue that balanced performance measurement helps managers to translate an organisation's often lofty vision and strategy statements into concrete terms; communicate organisational strategy and goals to employees; align employee actions by setting goals and linking them to performance targets; integrate strategic planning and budgeting, which usually occur independently and in parallel; and validate and adjust the chosen strategy. All this should be accomplished by applying – designing, communicating, reviewing, and acting on – a balanced set of measures that quantify organisational goals and strategy. However, this strategy-based view reveals very little about how managers actually use measures in day-to-day situations. The focus in literature is a top-management perspective on management processes and how performance management systems help to implement long-term strategy.

For a more detailed and day-to-day view of performance measurement system use Simons (2000) suggests that there are five broad uses for the information of a performance measurement system: decision making, control, signalling, education and learning, and external communication. In the category of decision making, managers acquire data to understand a decision's likely effects and evaluate its possible outcomes prior to making the decision, which may involve planning or coordination. Control use means that performance measurement information is used in a feedback loop in which performance data are fed back and compared with targets to see if they are being met; if necessary, corrective actions are taken to place the controlled object back under control. The aim of signalling use is to notify employees about what is important: by focusing on certain information, managers can signal priorities and goals to employees. With education and learning use, performance information can be used by the organisation to learn about the drivers and dynamics of the business, as well as about changes in the internal and external environments that might affect the organisation. The last use category, external communication, focuses on communicating the company's performance to its external stakeholders, such as investors, suppliers, and customers.

Simons' (2000) classification describes generic ways to use a performance measurement system, but it does not specify the situations in which managers use the measures. Focusing on the usage situations, Bourne *et al.* (2000) argue that regular meetings by managers to review the measures they are responsible for are needed. In their case study, all three case companies held monthly meetings that dealt with reviewing measures, discussing progress, and agreeing on necessary actions. In another case study, Bourne *et al.* (2005) report that the highest-performing business units of the case company held regular team meetings in which performance reports were reviewed, and performance was also discussed "at every opportunity", such as in one-to-one situations. Ukko *et al.* (2007) compared communication channels for performance measurement data in small- and medium-sized manufacturing companies and found that face-to-face communication, such as team or company meetings or discussions with foremen, contributes to the successful communication of measurement information relative to system-based communication, such as hand-outs, e-mail or intranet.

2.3 Use of performance measurement system in maintenance process

The dominant view in the literature on performance management in maintenance is strategic, taking a higher-level management perspective, where the assumed use of performance measurement is to pinpoint inefficiencies that line-managers can be held responsible for (Fitzgerald, 2007). Coetzee (1999) suggest this can be achieved in a holistic way by essentially focusing on four issues: downtime, number of tasks completed, proportion of time spent on scheduled tasks, and total maintenance cost. The argument is that such a set of performance measures provides higher-level management with the basis for decision making that takes the overall goals of the maintenance process in account, as well as affecting behaviour of the organisation in the field.

The proposal by Coetzee has been criticised as too simplistic by Kutucuoglu *et al.* (2001), who argued for the addition of customer perspective by including measures reflecting the impact of maintenance on the customer operations, such as number of late deliveries due to unscheduled downtime. Parida and Kumar (2006) emphasise that higher-level management are best served by maintenance performance measurement that balance external customer perspectives and internal efficiency measures Parida and Chattopadhyay (2007) argue for maintenance performance measurement with clearly articulated linkages between the measures in use and corporate strategy. Furthermore, in a maintenance process incorporating a number of different organisations the performance measurement systems of different organisations need to be aligned. In particular alignment is critical for a strategic view of the overall performance of maintenance process towards the final customer (Gaiardelli *et al.*, 2007).

Recently Muchiri *et al.* (2011) investigated the issue of leading and lagging measures in maintenance performance measurement. Their basic classification distinguishes between process and results, where maintenance process measures are leading indicators while measures of maintenance results are lagging indicators. Leading indicators are needed to determine the health of an asset and to monitor the effectiveness of maintenance process in real-time (Parida and Chattopadhyay, 2007). Lagging indicators are the basis for describing over-all performance and for deciding between maintenance, overhaul, or replacement. According to Muchiri *et al.* (2011) management needs to use a combination of leading and lagging indicators for assuring the sustained performance of a maintenance process.

2.4 Summary of literature review

In summary, existing general management literature contains some ideas about how performance measurements are actually used. The work of Simons (2000) highlights that there are different types of usage practices. Furthermore, the studies of Ukko *et al.* (2007) and Bourne *et al.* (2000) indicate that usage situations need to be taken into account when evaluating the effectiveness of performance measurement system use. However, this discussion on operational use is largely absent from maintenance management literature. The emerging, more operational level discussion focuses on the role of leading and lagging indicators in the design of the system. The literature does not yet elaborate the operational usage situations of performance measurement systems in use in the maintenance process. Thus, there is need for empirical research investigating the operational use of performance management systems in the maintenance process.

3. Methods

3.1 Research design

In order to discover and elaborate on actual practices of performance measurement system usage and their relationship with performance, we selected the inductive case study methodology for this research. Our literature review suggested that there are broad categories and guidelines on how performance measurement systems could be used but little information about specific usage practices and their relationships to performance. In this study, we employed the case study methodology to primarily understand the use of performance measurement systems as it emerges from the data. We take a theory building approach in which key variables and their linkages are identified (Voss *et al.*, 2002). The propositions of the conceptual model provided the theoretical foundation that guided our data collection on aspects of performance measurement systems other than their use.

The case organisation is the maintenance service unit of a global mechanical equipment company. The company manufactures, installs, and maintains large machines on every continent, with more than a thousand offices worldwide. The company's maintenance service unit offers field services for the equipment that the company and its primary competitors manufacture and install. Field service entails maintenance or repair services that are performed at the equipment's installation site by technicians who are sent from their home base to perform each maintenance or repair task. The case organisation has recently started to globally harmonise its field service operations processes. Part of the harmonisation involves the deployment of a performance measurement system for use by its country organisations and their field service units. The system has been deployed top-down by the global development organisation to local country organisations. Field service units include operational maintenance units that are geographically distributed and responsible for servicing the installed base of equipment in their geographical area. Each field service unit consists of a supervisor and the technicians under his or her management.

The system produces the same set of performance measures for four organisational levels: single field service technician, field service unit, service region, and country. The performance measurement system has thirteen measures on an electronic spreadsheet. Each measure is presented in a bar chart with monthly values and an optional trend line for the values. Measures are maintenance and service related operational measures such as times, numbers of instances, or percentage values or performance indices derived from them. There is only one measure with financial, monetary values. The measures are conventional in the context of maintenance process. The measures were selected by the organisation's global maintenance team. The objective of the designers was to form an overall picture of a technician's tasks and daily work routines. The implemented measures are backward looking by nature and based on historical data, whereas an expressed requirement from the organisation would have been for more forward looking measures to determine what should be done in order to achieve future goals.

The chosen case design is a single-case embedded design (Yin, 2003) in which the maintenance process of the company is the main unit of analysis and the field service units are the embedded units of analysis. We answer the study's research question primarily from the perspective of a field service unit because the bulk of activity in the use of the performance measurement system in this case takes place at the field unit level. However, we will also present findings on the global field service organisation level to summarise how the results apply to the tasks of the organisation's upper level management. Instead of a multiple case design, we chose the single-case embedded design because it helps to control for company-level and environmental influences on the constructs in the conceptual model. It is therefore easier to identify the effects of different usage practices on performance. The single-case focus also allowed us to study usage practices in operational detail.

3.2 Data collection approach: mixed methods

We collected the data using a mixed methods approach. The literature suggests that the use of both qualitative and quantitative data in case studies generates synergies by corroborating findings (Eisenhardt, 1989) and that the use of multiple sources of evidence triangulates data such that the construct validity of the study increases (Yin, 2003). We used three data collection methods:

(1) meetings with the case organisation's managers;

(2) interviews of case organisation employees; and

(3) a survey in the case organisation.

Our approach is a sequential exploratory approach in which qualitative data (from meetings and interviews) are collected and analysed first, followed by the collection and analysis of quantitative (survey) data. Finally these two types of data are synthesised (Creswell, 2009). Figure 2 illustrates the phases of the data collection and analysis processes.

3.3 Qualitative data: meetings and interviews

The meetings provided us with an understanding of the case organisation and helped with planning the interviews and the survey. The participants were a senior development manager and two other managers who were involved in the deployment of the performance measurement system. Each of the seven meetings lasted one to two hours. We took written notes on the discussed topics for subsequent analysis and for guidance in the research process. In addition, we held four follow-up meetings during or after the survey data collection and analysis to interpret, discuss, and validate the survey findings.

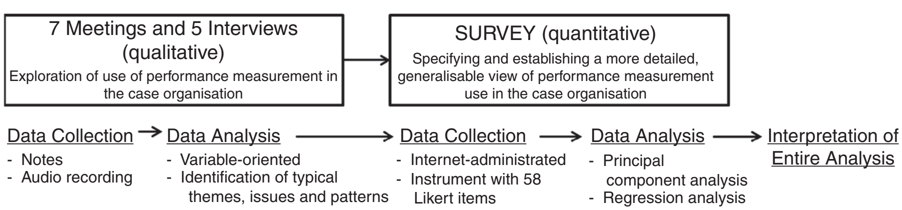
The purpose of the semi-structured interviews was to identify usage practices, usage situations, and purposes for performance measurement system use. The first author interviewed five field unit supervisors, who were the primary users of the performance measurement system. Each interview lasted approximately one hour. The interview themes were the operational context of the field service unit, general improvement and management methods, performance measurement system use, and the associated benefits and challenges perceived by the supervisors. Interviews were audio recorded and transcribed for analysis.

We analysed the qualitative data from the meetings and interviews through three concurrent activities:

(1) data reduction to separate the important pieces of data;

(2) data display to highlight and present reduced data in a cognitively easy form; and

(3) conclusion drawing and verification to explain what the data mean (Miles and Huberman, 1994).



Source: Framework by Creswell (2009)

Figure 2.
Sequential exploratory design of the study

The analysis was variable-oriented (Miles and Huberman, 1994) because the aim was to look for typical themes, issues, and patterns in the case organisation. As a result of the qualitative analysis, we identified the case organisation's performance measurement system usage practices. Drawing on the resulting understanding, we then selected and generated scales and single items for the survey instrument to be deployed on a large scale to determine performance relationships.

3.4 Quantitative data: survey to determine performance relationships

The role of the quantitative survey was to broaden our understanding of how supervisors in the case organisation use the performance measurement system. With the survey data from 149 supervisors, we expanded our study of the performance measurement system use practices that were identified through the qualitative analysis. As a new element, the survey data enabled the identification of performance relationships. Although a survey is usually considered to represent a research approach of its own rather than being a data collection method, it can be a viable part of data collection within a case study (Yin, 2003). One of the main benefits and reasons for choosing a survey as a data collection method is its efficiency: representative inferences can be made regarding a target population by surveying only a small fraction of the entire population (Dillman *et al.*, 2009). In this study, the survey data enabled us to evaluate the generalisability of the interview findings within the population of the entire case organisation.

The target population of the survey was composed of the case organisation's field service unit supervisors from the countries in which the performance measurement system was used. We drew a sample from these countries with the help of the case company managers. The sample included supervisors from the countries in which first, the performance measurement system was used on a field service unit level, second, the system had been deployed at least one year ago, and third, the information system implementation used for performance data collection were comparable. These sampling criteria ensured that the respondents were sufficiently knowledgeable about the surveyed topics and made the data from different countries comparable. The final sample comprised 236 supervisors from seven countries (nine to 69 per country). In total, 149 responses were received (seven to 39 per country), resulting in a response rate of 63 per cent (43-73 per cent per country). The responses were collected with an internet-administered survey instrument. Supervisors were sent an e-mail invitation asking them to take part in the survey and received a follow-up e-mail one week later from their country managers. We tested for non-response bias by comparing early and later respondents, as suggested by Armstrong and Overton (1977). The responses to the main empirical variables did not differ statistically significantly between the respondents who answered after the first invitation and those that responded after the follow-up ($p < 0.05$ with independent samples t-tests).

We developed the survey instrument with the help of the extant literature, advice from experts (meetings with the case organisation managers), and interviews with representatives from the survey population (supervisors). This approach is recommended (Rossi *et al.*, 1983) and typical in operations management (Hensley, 1999). The constructs in the conceptual model guided the development of the survey instrument. We operationalised each construct with multiple items and developed the items for the Deployment of the Performance Measurement System, Behaviour, and Performance constructs using the literature and existing scales when possible. The items for performance measurement system use are based on the usage practices identified in the qualitative analysis. We categorised the items within each construct according different topics to obtain guidance for further analysis. Topics for the Deployment of the Performance Measurement System construct were chosen from the categorisation of Johnston and Clark (2008, p. 358). For the Use of the Performance Measurement System construct, we employ the performance measurement system use categorisation of Simons (2000, p. 67) and for the Behaviour construct, we use the performance measurement purpose categorisation of Johnston and Clark (2008, p. 359). In order to ensure that the supervisors understood the questions, we and the case organisation managers adapted the content and wording of each item to the case organisation. We conducted two face-to-face pre-tests with supervisors to ensure the unambiguity and understandability of the instrument. Furthermore, each survey was translated into the supervisors' native languages by the authors, the case company managers, and our researcher colleagues. All the survey items, initial topics, and citations, where appropriate, are listed in the Appendix.

The analysis of the survey data progressed in two consecutive steps. First, we sought to elaborate on the conceptual model by identifying the dimensions of the broad constructs in the conceptual model. We used principal component analysis, with procedures suggested by Hair *et al.* (2010), for this purpose. We could use 119-139 responses out of 149 in the analyses. We used the pairwise method for treating missing data to maximise the use of valid data because Little's MCAR tests resulted in insignificant p-values (>0.05). We ran a principal component analysis for each of the four constructs in the conceptual model. We determined the number of components using a combination of latent root criterion 1, scree test criterion, comparison with the a priori categorisation (see the Appendix), the percentages of variance extracted and commonalities. We excluded items with high cross-loadings (multiple loadings over 0.50) when the ideas of the items to be excluded were sufficiently accounted for by other items. We also checked that the remaining items in each factor still formed a coherent whole and reran the principal component analysis. Finally, we calculated summated scales from the solutions based on loadings higher than 0.50. Hair *et al.* (2010) suggest that with a sample size of 120, a significance level of 0.05 and a power level of 80 per cent, loadings over 0.50 are statistically significant and that loadings with absolute values of 0.50 or greater can be considered practically significant.

In the second analysis step, we examined the propositions of the conceptual model by using the summated scales from the first step. We employed a stepwise multiple regression procedure (Hair *et al.*, 2010). The purpose of the regression analysis was twofold:

(1) to determine the sets of independent variables that best explain the dependent variables; and

(2) to examine the relative importance of independent variables in the estimated regression equations.

We estimated the models using a variable inclusion criterion of $p = 0.01$ and exclusion criterion of $p = 0.011$. We accommodated missing data with pairwise deletion, as Little's MCAR test, with a significance of $p = 0.062$, indicated that data were missing completely at random. We visually screened for influential observations by using scatterplots of residuals and scatterplots between dependent and independent variables. All models were eventually estimated with all the data because the removal of a few possible influential observations resulted only in minor changes in the models. Finally, we assessed multicollinearity by examining the tolerance values of the variables. The final outcome of the regression analysis was a single regression model for each dependent variable.

3.5 Validity

According to Yin (2003), the criteria for judging the quality of a case study design comprise four tests: construct validity, internal validity, external validity, and reliability. Because this study did not include causal research, the evaluation of internal validity can be omitted. We will discuss the three other tests briefly.

To enhance construct validity, we used multiple sources of evidence: meetings, interviews, and the survey. We used data from meetings and interviews to identify performance measurement system usage practices in the case organisation and further refined and verified these results with the survey data. This sequential approach ensured that the survey focused on the right issues because its construction relied on the other data collection methods. We find the construct validity of the study to be satisfactory overall because all the broad constructs of the conceptual model were operationalised with summated scales derived through a structured approach that used existing theory, qualitative research and refinement with the survey. Moreover, the nomological validity of the summated scales can be assumed because their relations were mainly in line with *P1-P3*.

The external validity, or generalisability, of this study stems from the selection of a case organisation that is representative of a typical field service organisation offering preventive maintenance and repair services. Its performance measurement system consists of measures that are typical for maintenance processes. Using a typical, representative case in a single case study is advised by Yin (2003). The potential boundary conditions to our findings stemming from the case company characteristics include its geographically distributed organisational form, small field service units, average technological complexity of maintained equipment, and mature stage of the business.

Finally, we addressed reliability through a thorough documentation of the research design and data collection procedures, as suggested by Yin (2003). Meeting topics and agendas were designed in advance and documented, meeting notes were taken, interviews were semi-structured and audio-recorded, the construction of the survey instrument was documented, sand the survey data analysis was described in detail. Thus, we believe that another researcher could perform the same case study again, yielding similar results.

4. Results and analysis

This section presents the results derived from the analysis of the survey data. Section 4.1 introduces the summated scales established through principal component analysis and discusses their meaning. Section 4.2 then presents the results of the regression analyses, which explore and elaborate on the propositions of the conceptual model. Finally, Section 4.3 synthesises the results and outlines the main findings of the study.

4.1 Summated scales

Factor analysis of the data capturing the conceptual model's four broad constructs resulted in three factors for each broad construct, yielding a total of 12. We interpreted each factor together with the case company managers and with the help of the conceptual model. Next, we calculated summated scales for use in the regression analysis. Table I summarises the names and descriptions of the scales. The Appendix contains the items of each scale and their loadings.

As suggested by Hair *et al.* (2010), we assessed the scales for reliability (Cronbach's α values), unidimensionality (loadings in principal component analysis), and content (conceptual model and discussions with case organisation managers) and discriminant validity (correlations). Convergent validity could not be assessed due to the lack of concepts available for comparison, and nomological validity was considered to be satisfactory, as stated in Section 3.5. The only perceived problems with validity were the somewhat low content validity of the Improve Competitiveness scale, the low discriminant validity between the Inspect and Improve and the Motivate scales (correlation of 0.787), and the low reliability of the Reporting scale ($\alpha = 0.541$) compared to the suggested threshold of 0.70 (Hair *et al.*, 2010; Hinkin, 1995). However, taking the explorative nature of the scale creation process into consideration, we are satisfied with the scales' ability to capture the relevant aspects of performance measurement in the case organisation and that they can be used to analyse the relationships suggested in propositions of the conceptual model. In particular, our close collaboration with the case company's managers and the inductive approach used in the creation of the items increases our confidence. Table II presents the descriptive statistics and the bivariate correlations of the summated scales.

4.2 Exploring the propositions of the conceptual model

We studied and elaborated on the relationships suggested by *P1-P3* using stepwise regression analysis. Nine summated scales of the three latter constructs of the conceptual model were chosen as dependent variables and a multivariate regression

Name	Description
Construct: deployment of performance measurement system	
Proper system design	How well the performance measurement system is designed: if it is applicable, understandable and appropriate for use in management
Training	How well the supervisors have been trained and advised to use the performance measurement system
Reporting	How accessible and easily available the performance measurement system's report is
Construct: use of performance measurement system	
Inspect and Improve	Performance measurement system use in which data are first analysed and inspected; thereafter, improvement actions are undertaken
Decision Making	Performance measurement system use in which managerial, field service unit level decisions are made
Motivate	Personal level, motivational use of performance measurement system
Construct: behaviour	
Improve competitiveness	Willingness to act and improve operations in a way that is beneficial for business
Motivation	Motivation of the supervisor and technicians to perform well in their operations
Goal communication	Supervisor's and technicians' awareness of the goals and strategies of the organisation
Construct: performance	
Efficiency	Performance related to inputs and outputs of the operations
Dedication	Performance related to effectiveness, pride and devotion towards work
Extra work	Performance related to proactivity and entrepreneurial drive

Table I. Names and descriptions of the summated scales

Table II. Correlations and descriptive statistics of the summated scales

Summated scale	1	2	3	4	5	6	7	8	9	10	11	12
1. Proper system design	1											
2. Training	0.539**	1										
3. Reporting	0.516**	0.429**	1									
4. Inspect and Improve	0.519**	0.312**	0.385**	1								
5. Decision Making	0.458**	0.283**	0.209*	0.691**	1							
6. Motivate	0.450**	0.152	0.275**	0.787**	0.614**	1						
7. Improve competitiveness	0.440**	0.316**	0.290**	0.328**	0.238*	0.358**	1					
8. Motivation	0.331**	0.221*	0.088	0.297**	0.242**	0.371**	0.671**	1				
9. Communication	0.540**	0.273**	0.341**	0.335**	0.180	0.318**	0.623**	0.633**	1			
10. Efficiency	0.212*	0.240**	0.043	0.084	0.022	0.141	0.424**	0.436**	0.230*	1		
11. Dedication	0.074	0.095	−0.125	0.038	0.065	0.043	0.257**	0.542**	0.180	0.558**	1	
12. Extra work	0.006	−0.033	−0.106	−0.066	−0.136	0.008	−0.007	0.159	0.142	0.389**	0.299**	1
Mean	4.56	3.86	4.93	3.44	2.64	3.16	5.48	5.50	5.59	5.14	5.32	4.82
SD	0.111	0.153	0.122	0.092	0.099	0.097	0.080	0.079	0.075	0.074	0.078	0.119
n (for Mean and SD)	111	119	119	117	117	119	136	139	137	133	136	136

Notes: *,**Correlation is significant at the 0.05 and 0.01 level, respectively (two-tailed) model was estimated for each them. Independent variables added to the stepwise model estimation process were summated scales from the preceding constructs and 12 conceptual variables that we included to control for influences that were external to the conceptual model. These potential external influences were identified in the interviews with the supervisors and meetings with the case company managers. The control variables included the characteristics of the field service units' personnel (supervisor's years of experience, number or technicians), installed base (equipment use, variety of equipment), and customers (business type, contract type), as well as country-level differences absorbed by the dichotomous country variables. We acquired the data for the control variables from the case company's enterprise resource planning software. In the resulting models, the independent variables to be entered for each model were chosen by the stepwise estimation technique. Table III presents the results of the regression analyses.

We examined the residual and normal probability plots of the residuals for the nine variates and assessed that the assumptions of the regression analysis hold adequately for the purposes of our analysis. Multicollinearity is a small concern in the estimation of models for two dependent variables, Motivation and Efficiency, even though the tolerance values are above the reference point of 0.33 (Hair *et al.*, 2010). As a dependent variable, Motivation may also be explained by the independent variable Inspect and Improve, which has a high correlation (0.787) with the independent variable that is in the model, Motivate. In the model for the dependent variable Efficiency, which is explained by the independent variable Motivation, a possible explanatory variable could be Improve Competitiveness, which has a correlation with Efficiency (0.424) that is of comparable size to the correlation between Efficiency and Motivation (0.436). We address these possible effects as well in the interpretation of the results in the next sections.

4.3 Summary of findings: usage practices and their contribution to performance

Given the research question, the first result of this study is the identification of three performance measurement system usage practices: Inspect and Improve, Decision Making, and Motivate. In the case company, Decision Making clearly differs from the two others because it captures the higher-level management of the supervisors in making decisions related to an entire organisational unit (field service unit). Usage practices Inspect and Improve and Motivate, on the other hand, revolve around people (technicians) and their management. Inspect and Improve is a control-type of usage practice in which control action is tied to personal level improvement whereas Motivate is a more leadership-type usage practice. We summarise this categorisation of the usage practices as the first finding:

- Finding 1: three usage practices are relevant when managing with a performance measurement system: Inspect and Improve, Decision Making, and Motivate.

The regression analyses suggest relationships between identified usage practices and performance as summarised in Figure 3. These relationships extend the conceptual model and *P1*-*P3*.

P1 suggests a positive link between the deployment of a performance measurement system and its use. Proper System Design is positively related to all the three usage practices of Inspect and Improve, Decision Making, and Motivate (β coefficients are 0.517, 0.463, and 0.480, respectively). This empirical finding supports *P1*. However, the use of performance measurement systems varies among countries, as demonstrated by four β coefficients between usage practices and country variables

Table III. Estimated regression models for the nine dependent variables

Independent variables	Inspect and Improve	Decision Making	Motivate	Improve Competitiveness	Motivation	Goal Communication	Efficiency	Dedication	Extra Work
	Dependent variables[c]								
Constant[a]	1.626**	0.863***	0.982***	3.805**	4.418**	3.666**	3.059**	2.363**	4.821**
Proper System Design	0.517**	0.463**	0.480**	0.440**		0.540**			
Motivate					0.371**				
Motivation							0.436**	0.542**	
NL[b]	−0.278**	−0.368**							
UK[b]	−0.233**								
Spain[b]			0.273*						
Sweden[b]									−0.282*
Summary data of the models for each dependent variable									
n (pairwise, smallest)	101	101	101	101	101	101	101	101	101
R^2	0.382	0.345	0.276	0.194	0.138	0.291	0.190	0.294	0.080
Adjusted R^2	0.360	0.330	0.259	0.184	0.128	0.283	0.181	0.286	0.069
df (ANOVA)	87	87	87	87	87	87	87	87	87
F	17.339**	22.400**	16.203**	20.676**	13.764**	35.372**	20.213**	35.833**	7.445*
Smallest tolerance and related variable	0.547 Contract type[d]	0.560 Training	0.708 Training	0.710 Training	0.380 Inspect and Improve	0.710 Training	0.550 Improve competitiveness	0.550 Improve competitiveness	0.901 Equipment use[d]

Notes: NL, the Netherlands; UK, the United Kingdom. [a]Coefficients of the variable are unstandardised. [b]Dummy coded variable for binary data. [c]Cells between dependent and independent variables are standardised β coefficients. [d]Contextual variable. $*p \leqslant 0.01$; $**p \leqslant 0.001$; $***p \leqslant 0.05$

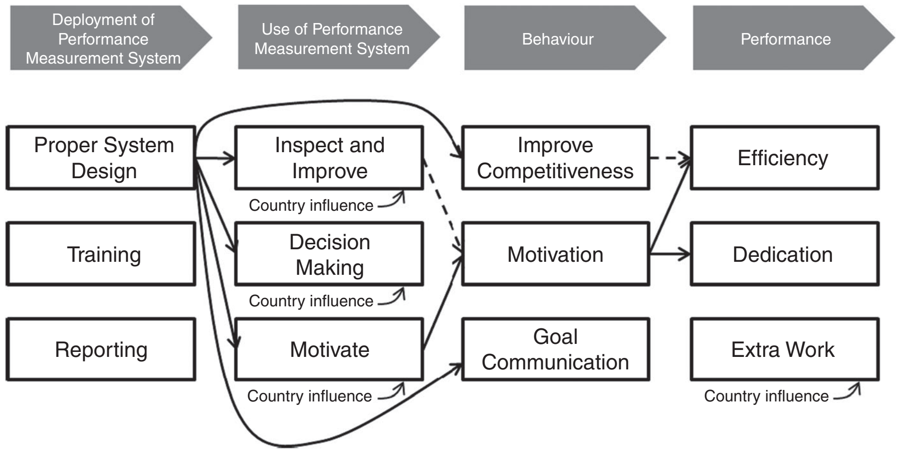


Figure 3. Relationships suggested by regression analyses (solid lines) and correlations (dotted lines)

(−0.368, −0.278, −0.233, 0.273). Nevertheless, Proper System Design has a larger effect, as seen from the β coefficients. Therefore, we conclude:

- Finding 2: the proper design of a performance measurement system is positively related to its use.

P2 suggests that the use of a performance measurement system stimulates beneficial behaviour. As expected, Motivation is positively related to the Motivate use scale (β coefficient 0.371). However, Proper System Design explains the behavioural scales of Improve Competitiveness and Goal Communication (β coefficients of 0.440 and 0.540, respectively). Moreover, the Inspect and Improve usage practice can be interpreted to be another contributor to Motivation due to their high bivariate correlation (Table II). In light of *P2*, our mixed findings suggest that a set of different relationships explain behavioural effects. We summarise these relationships as Findings 3-5:

- Finding 3: the use of a performance measurement system for personal-level motivational and improvement practices is positively related to motivation.
- Finding 4: the proper design of a performance measurement system is positively related to the willingness to improve competitiveness.
- Finding 5: the proper design of a performance measurement system is positively related to awareness of organisational goals.

P3 suggests that beneficial behaviour improves organisational performance. According to our analysis, the behaviour of Motivation is positively associated with the performance areas of Efficiency and Dedication (β coefficients of 0.436 and 0.542, respectively). In addition, the high bivariate correlation suggests that the Improve Competitiveness behaviour has a positive effect on Efficiency. To our surprise, the performance scale of Extra Work was not explained by any behaviours but by a contextual country variable. Further discussions with case company managers clarified that this result reflects country-level differences in the processes related to extra work. Nonetheless, *P3* seems to be valid in light of these findings. We summarise this in Findings 6 and 7:

- Finding 6: motivation and willingness to improve competitiveness are positively related to organisational efficiency.
- Finding 7: motivation is positively related to performance area dedication.

Finally, some of the scales were not included as independent variables in any of the estimated regression models using the stepwise algorithm. In contrast to the suggestions of *P1-P3*, Training and Reporting did not explain the use of a performance measurement system, the Decision Making usage practice did not explain behaviour, and Goal Communication did not explain performance. From our perspective, this result is part of the elaboration of the conceptual model because it emphasises the other variables that have effects. A summary of the Findings and how they support or discredit *P1-P3* are listed in Table IV.

5. Discussion

The conducted study provides a complementary, operational-level view of performance measurement to the strategic, higher-level management perspective provided by current literature on maintenance performance management (Coetzee, 1999; Kutucuoglu *et al.*, 2001; Parida and Kumar, 2006; Gaiardelli *et al.*, 2007; Muchiri *et al.*, 2011). The study identifies a set of situations to be investigated in the evaluation of outcomes on operational-level from introducing a performance measurement system in a maintenance process. In the section that follows we discuss the reasons for the observed effects of the performance management system in the case organisation, the role of organisational context, and practical implications. A note on limitations concludes the discussion.

5.1 Implications for the constitution of the performance effects of performance measurement

In this study, we elaborate the conceptual model (*P1-P3*) derived from literature and through the elaboration develop a more detailed operational perspective on the performance effects of performance measurement systems. Overall, our findings are in alignment with the extant literature on how performance measurement affects organisational performance as captured through *P1-P3*. According to our analysis (see Table IV), all three propositions hold individually: Finding 2 supports *P1*, Finding 3 supports *P2*, and Findings 6 and 7 support *P3*. Findings 2, 3, 6, and 7 are also in line with the overall conceptual model (Figure 1), which posits that performance management practices influence behavioural effects, which subsequently influence

Finding	Relation	Proposition
1: Three usage practices are relevant when managing with a performance measurement system: Inspect and Improve, Decision Making, and Motivate	–	–
2: Proper design of a performance measurement system is positively related to its use	Supports	*P1*
3: The use of performance measurement system for personal level motivational and improvement practices is positively related to motivation	Supports	*P2*
4: The proper design of a performance measurement system is positively related to willingness to improve competitiveness	Discredits	*P2*
5: The proper design of a performance measurement system is positively related to awareness of organisational goals	Discredits	*P2*
6: Motivation and willingness to improve competitiveness are positively related to organisational efficiency	Supports	*P3*
7: Motivation is positively related to performance area dedication	Supports	*P3*

Table IV. Findings of the study and their relations with the study's propositions performance (de Leeuw and van den Berg, 2011). Unexpectedly, Findings 4 and 5 suggest that for Goal Communication and Improve Competitiveness, the design of a performance measurement system matters more than its actual use, which implies that the constitution of the performance effect of performance measurement does not necessarily follow a single path, as suggested by the conceptual model.

The first implication of our elaborated model is to highlight the role of motivation. The behavioural effect of Motivation links the use of a performance measurement system and performance (Findings 3, 6, and 7). More specifically, Finding 3 suggests that the motivational (Motivate) and personal-level improvement (Inspect and Improve) uses are the most effective practices for increasing Motivation. Furthermore, Motivation stands out as the clearest antecedent to performance (Findings 6 and 7). In the existing literature, motivation is discussed but not extensively emphasised. For instance, Johnston and Clark (2008) identify motivation as one of four purposes of performance measurement, and Simons (2000) states that performance measurement system use for control induces motivation – both intrinsic and extrinsic – in an organisation. However, given our results, future studies on the performance effects of performance measurement are well advised to consider motivation as a major source of increased performance. We argue that intrinsic motivation in particular is beneficial in improving performance, in contrast to extrinsic motivation induced by tangible rewards.

Second, our detailed analysis demonstrates both expected and unexpected effects of proper performance measurement system design. In our analysis, Proper System Design captures user views of the performance measurement system. Positive evaluations of Proper System Design are associated with higher levels of performance measurement system use (Finding 2), as expected (Franco and Bourne, 2003; de Waal, 2003). This further supports previous findings of performance measurement system implementation success. However, unlike what was expected (*P2*), Findings 4 and 5 indicate a direct relationship between Proper System Design and the behavioural effects of Goal Communication and Improving Competitiveness. Due to the lack of specific usage practice interventions, we label these mechanisms as "informational effects". As the main theoretical basis for *P2*, de Leeuw and van den Berg (2011) found somewhat similar informational effects in their study. However, their behavioural clusters of Understanding and Motivation relate rather directly to performance indicators, whereas our findings suggest that induced behaviour is more general and all-purpose.

In summary, our explorative analysis contributes to the understanding of the operational constitution of the performance effects of performance measurement systems. First, the use of a performance measurement system to motivate and the resulting motivation effect was a central mechanism in producing performance outcomes in the studied case. Second, our findings on the direct behavioural effects of Proper System Design suggest that informational effects are a distinct type of mechanism that does not require use. Instead the design of the system itself informs – without necessarily being used – what are required behavioural effects. However, these findings describe the case organisation and its 149 local units, and it is therefore possible that other organisations would exhibit different mechanisms. In order to establish the foundation for generalisation, we continue with a discussion of the relationship between organisational context and the use of performance measurement systems.

5.2 Organisational context and the use of performance measurement systems

Our exploratory findings on the use of performance measurement systems (Finding 1) differ from the model of Simons (2000). Simons lists five uses for a performance measurement system: Control, Signalling, Decision making, Education and learning, and External communication. Our principal component analysis suggests that three dimensions of use constitute the best model in this case: Inspect and Improve, Decision Making, and Motivate. Although External communication was not examined in this study, the other dimensions demonstrate similarities. Most notably, the Decision Making use is part of both models. The Inspect and Improve practice overlaps both the Control use and the Education and learning use because its control action is linked to the improvement and coaching of the technicians. The Motivate practice is another distinct component of the Control use. Finally, Signalling is partly included in the Inspect and Improve practice because improvement in this area is related to communicating and signalling improvement areas to people. Overall, our analysis produced a different categorisation for the use of performance measurement systems, although some similarities could be found.

Our findings emphasise a softer management style in the effective use of performance measurement systems. Instead of upper level management and control it is motivation and leadership interaction with employees that are central for achieving desired performance outcomes (see Figure 3). Marr's (2004) survey findings, in contrast to our findings, identify motivation as a less-frequent reason for using a performance measurement system compared to control and strategy. Furthermore, the descriptive statistics (Table II) show that supervisors rate their frequency of use for Motivate and Inspect and Improve higher than for Decision Making. From our perspective, this reflects the case organisation's successes in training field service unit supervisors in a coaching style of management, as was also revealed by the managers in the meetings. There has been a development programme, for instance, in which the supervisors have been trained to act as coaches for the field technicians and to help them to develop technical expertise. Ties between management style and the use of performance measurement systems have also been suggested by Bititci *et al.* (2006). In the case organisation, the people-centred use of the performance measurement system has a positive relation with the organisation's efforts in coaching and personal development.

This implies a contingent relationship between the use of a performance measurement system and the management style of the organisation. In their multiple-case study, Bititci *et al.* (2006) suggest that successful performance measurement system implementation requires an authoritative management style and that the subsequent success leads to an achievement-oriented culture and a more consultative and participative management style. Our findings (Finding 1) provide a more detailed description of the specific types of usage practices that are associated with performance in organisations that are managed in a consultative and participative manner. To be clear, our study does not elaborate the more traditional use of performance measurement systems as a tool for mechanistic control. However, we propose further research into understanding alternative ways of using performance measurement systems and how their effectiveness is contingent on the organisational context.

5.3 Practical implications

The findings of this study offer insights on how managers can use performance measurement systems to improve performance, based on a case example from maintenance. To position the study's practical implications, we use the design proposition framework of Denyer *et al.* (2008), which identifies four elements: context, intervention, mechanism, and outcome (Denyer *et al.*, 2008). Mechanism (motivation) and outcome (improved performance) have already been extensively discussed in the preceding section. Here, in discussing practical implications, we focus on the first two elements of the framework: context and intervention.

Regarding the context issue in relation to performance measurement, our study suggests that a performance measurement system can successfully be used as a motivational leadership tool to influence the working habits of the field services personnel. In the case company, supervisors used the performance measurement system to motivate field service technicians and to identify opportunities to achieve personal improvement. Similar usage practices are likely to be applicable in many contexts in which the maintenance organisation consists of many small units of a manager and a small team of technicians.

Regarding the performance measurement intervention, our analysis suggests that the design and implementation of the performance measurement system are central antecedents to its use. This is a reminder for higher-level managers who are responsible for the design and deployment of performance measurement systems. It is critical for management to ensure that systems are actually taken into use in the organisation. When systems are not used, it is the responsibility of management to investigate possible changes in the design of the system, the implementation, and training procedures facilitate wider use. The importance of proper design and implementation is further underlined by the informational mechanisms that can stimulate beneficial behaviour when the performance measurement system is available, but not necessarily taken into use.

In addition to design and implementation, the higher-level managers who are responsible for the deployment of the performance measurement system should pay attention to differences among the organisational units that use the system. In this study, we found that preferred usage practices varied among country organisations. Although "country" was addressed only as a control variable in the study, the country managers' effort and commitment towards the system have a significant influence on its use. In meetings, managers of the global-level organisation told us that some country managers are very enthusiastic about the system, whereas there are others who pay little attention to it. The importance of management leadership, commitment and desire for a performance measurement system has also been identified as a factor that enables performance management in previous studies (Franco and Bourne, 2003; Nudurupati and Bititci, 2005).

5.4 Limitations

We are generally satisfied with the validity of our case study, as discussed in Section 3.5. As with any case study, our findings would benefit from further research in different settings. The primary limitation of this study is its reliance on meetings and interviews for the identification of usage practices. If some relevant performance measurement practice was not brought up in the meetings or interviews, it would likely not surface in the survey. However, the expertise of the case company's senior development manager and two other managers, and including open questions in the survey instrument, mitigate against this risk.

The results of the survey also had a few minor limitations, as briefly reviewed in results Sections 4.1 and 4.2. One of the created summated scales (Reporting, $\alpha = 0.541$) had an α value below the suggested threshold of 0.7 and another was marginally reliable (Dedication, $\alpha = 0.699$). The convergent validity of the scales could not be assessed due to lack of similar measures, and discriminant validity between the Inspect and Improve and Motivate scales was rather low. These problems were caused by the exploratory nature of the survey: scale creation took place during the survey and it was more exploratory than what is traditionally done to address a sufficiently large field of performance measurement topics. Nonetheless, we believe that the summated scales are sufficiently valid and reliable to serve their exploratory purpose. In future research, their items should be revised, which would also address some minor issues with multicollinearity and thus produce more accurate estimates.

6. Conclusions

This paper has discussed different usage practices of performance measurement systems and their role in the constitution of the performance effects of performance measurement. Prior literature has identified broad categories of performance measurement system use and focused rather heavily on strategic, upper-level management in the use of these systems. In this exploratory single-case study, we examine usage practices and their contribution to improved performance in more detail in maintenance. According to our analysis, the main usage practices of performance measurement system are Inspect and Improve, Decision Making, and Motivate. Regarding the performance effects, Inspect and Improve and Motivate were the central usage practices contributing to improved performance in the maintenance process. Furthermore, our findings emphasise motivation as an important behavioural factor that is achieved with the use of performance measurement systems and that relates to improved performance.

There are future research opportunities in building on our findings to expand the understanding of performance measurement system use. In particular, motivational, leadership use and the motivation induced with use have been dominated by the strategic control views in current literature (Coetzee, 1999; Kutucuoglu *et al.*, 2001; Parida and Kumar, 2006; Gaiardelli *et al.*, 2007; Muchiri *et al.*, 2011). Our findings suggest that these areas are important to capture the potential of performance measurement system use for some organisations. Another direction for further research is the beneficial performance effects that stem from the deployment of a performance measurement system rather than its use. We identified these influences as an informational mechanism. Future research is needed to specify how the emergent understanding of the operational use of performance measurement systems can contribute to the body of knowledge on the design and implementation of performance measurement systems in maintenance processes.

References

Armstrong, J.S. and Overton, T.S. (1977), "Estimating nonresponse bias in mail surveys", *Journal of Marketing Research*, Vol. 14, pp. 396-402.

Bititci, U.S., Carrie, A.S. and McDevitt, L. (1997), "Integrated performance measurement systems: a development guide", *International Journal of Operations & Production Management*, Vol. 17 No. 5, pp. 522-534.

Bititci, U.S., Mendibil, K., Nudurupati, S., Garengo, P. and Turner, T. (2006), "Dynamics of performance measurement and organisational culture", *International Journal of Operations & Production Management*, Vol. 26 No. 12, pp. 1325-1350.

Bourne, M., Kennerley, M. and Franco-Santos, M. (2005), "Managing through measures: a study of impact on performance", *Journal of Manufacturing Technology Management*, Vol. 16 No. 4, pp. 373-395.

Bourne, M., Mills, J., Wilcox, M., Neely, A. and Platts, K. (2000), "Designing, implementing and updating performance measurement systems", *International Journal of Operations & Production Management*, Vol. 20 No. 7, pp. 754-771.

Coetzee, J.L. (1999), "A holistic approach to the maintenance 'problem'", *Journal of Quality in Maintenance Engineering*, Vol. 5 No. 3, pp. 276-281.

Creswell, J.W. (2009), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd ed., Sage, Los Angeles, CA.

Dean, J.W. and Snell, S.A. (1991), "Integrated manufacturing and job design: moderating effects of organizational inertia", *The Academy of Management Journal*, Vol. 34 No. 4, pp. 776-804.

de Leeuw, S. and van den Berg, J.P. (2011), "Improving operational performance by influencing shopfloor behavior via performance management practices", *Journal of Operations Management*, Vol. 29 No. 3, pp. 224-235.

Denyer, D., Tranfield, D. and van Aken, J.E. (2008), "Developing design propositions through research synthesis", *Organization Studies*, Vol. 29 No. 3, pp. 393-413.

de Waal, A.A. (2003), "Behavioral factors important for the successful implementation and use of performance management systems", *Management Decision*, Vol. 41 No. 8, pp. 688-697.

Dillman, D.A., Smyth, J.D. and Christian, L.M. (2009), *Internet, Mail, and Mixed-Mode Surveys: the Tailored Design Method*, 3rd ed., Wiley, Hoboken, NJ.

Douglas, T.J. and Fredendall, L.D. (2004), "Evaluating the deming management model of total quality in services", *Decision Sciences*, Vol. 35 No. 3, pp. 393-422.

Eisenhardt, K.M. (1989), "Building theories from case study research", *The Academy of Management Review*, Vol. 14 No. 4, pp. 532-550.

Fitzgerald, L. (2007), "Performance measurement", in Hopper, T., Northcott, D. and Scapens, R.W. (Eds), *Issues in Management Accounting*, Pearson Education, pp. 223-241.

Flynn, B.B. and Flynn, E.J. (2004), "An exploratory study of the nature of cumulative capabilities", *Journal of Operations Management*, Vol. 22 No. 5, pp. 439-457.

Franco, M. and Bourne, M. (2003), "Factors that play a role in 'managing through measures' ", *Management Decision*, Vol. 41 No. 8, pp. 698-710.

Franco-Santos, M., Kennerley, M., Micheli, P., Martinez, V., Mason, S., Marr, B., Gray, D. and Neeley, A. (2007), "Towards a definition of a business performance measurement system", *International Journal of Operations & Production Management*, Vol. 27 No. 8, pp. 784-801.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010), *Multivariate Data Analysis: A Global Perspective*, 7th ed., Prentice Hall, Upper Saddle River, NJ.

Hensley, R.L. (1999), "A review of operations management studies using scale development techniques", *Journal of Operations Management*, Vol. 17 No. 3, pp. 343-358.

Hinkin, T.R. (1995), "A review of scale development practices in the study of organizations", *Journal of Management*, Vol. 21 No. 5, pp. 967-988.

Gaiardelli, P., Saccani, N. and Songini, L. (2007), "Performance measurement systems in after-sales service: an integrated framework", *International Journal of Business Performance Management*, Vol. 9 No. 2, pp. 145-171.

Ittner, C.D. and Larcker, D.F. (2003), "Coming up short on nonfinancial performance measurement", *Harvard Business Review*, Vol. 81 No. 11, pp. 88-95.

Jambulingam, T., Kathuria, R. and Doucette, W.R. (2005), "Entrepreneurial orientation as a basis for classification within a service industry: the case of retail pharmacy industry", *Journal of Operations Management*, Vol. 23 No. 1, pp. 23-42.

Johnston, R. and Clark, G. (2008), *Service Operations Management: Improving Service Delivery*, 3rd ed., FT Prentice Hall, Harlow.

Kaplan, R.S. and Norton, D.P. (1992), "The balanced scorecard – measures that drive performance", *Harvard Business Review*, Vol. 70 No. 1, pp. 71-79.

Kaplan, R.S. and Norton, D.P. (1996), "Using the balanced scorecard as a strategic management system", *Harvard Business Review*, Vol. 74 No. 1, pp. 75-85.

Kutucuoglu, K.Y., Hamali, J., Irani, Z. and Sharp, J.M. (2001), "A framework for managing maintenance using performance measurement systems", *International Journal of Operations & Production Management*, Vol. 21 Nos 1/2, pp. 173-195.

Marr, B. (2004), *Business Performance Management: Current State of the Art. A Survey Report*, Hyperion Solutions; Cranfield School of Management, Cranfield, available at: http://dspace.lib.cranfield.ac.uk/handle/1826/1222 (accessed 12 November 2012).

Melnyk, S.A., Stewart, D.M. and Swink, M. (2004), "Metrics and performance measurement in operations management: dealing with the metrics maze", *Journal of Operations Management*, Vol. 22 No. 3, pp. 209-218.

Miles, M.B. and Huberman, A.M. (1994), *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd ed., Sage, Thousand Oaks, CA.

Muchiri, P., Pintelon, L., Gelders, L. and Martin, H. (2011), "Development of maintenance function performance measurement framework and indicators", *International Journal of Production Economics*, Vol. 131 No. 1, pp. 295-302.

Neely, A. and Bourne, M. (2000), "Why measurement initiatives fail", *Measuring Business Excellence*, Vol. 4 No. 4, pp. 3-6.

Neely, A., Gregory, M. and Platts, K. (1995), "Performance measurement system design: a literature review and research agenda", *International Journal of Operations & Production Management*, Vol. 15 No. 4, pp. 80-116.

Neely, A., Gregory, M. and Platts, K. (2005), "Performance measurement system design: a literature review and research agenda", *International Journal of Operations & Production Management*, Vol. 25 No. 12, pp. 1228-1263.

Neely, A., Richards, H., Mills, J., Platts, K. and Bourne, M. (1997), "Designing performance measures: a structured approach", *International Journal of Operations & Production Management*, Vol. 17 No. 11, pp. 1131-1152.

Nudurupati, S.S. and Bititci, U.S. (2005), "Implementation and impact of IT-supported performance measurement systems", *Production Planning & Control*, Vol. 16 No. 2, pp. 152-162.

Parida, A. and Chattopadhyay, G. (2007), "Development of a multi-criteria hierarchical framework for maintenance performance measurement (MPM)", *Journal of Quality in Maintenance Engineering*, Vol. 13 No. 3, pp. 241-258.

Parida, A. and Kumar, U. (2006), "Maintenance performance measurement (MPM): issues and challenges", *Journal of Quality in Maintenance Engineering*, Vol. 12 No. 3, pp. 239-251.

Pavlov, A. and Bourne, M. (2011), "Explaining the effects of performance measurement on performance: an organizational routines perspective", *International Journal of Operations & Production Management*, Vol. 31 No. 1, pp. 101-122.

Radnor, Z.J. and Barnes, D. (2007), "Historical analysis of performance measurement and management in operations management", *International Journal of Productivity and Performance Management*, Vol. 56 Nos 5/6, pp. 384-396.

Rossi, P.H., Wright, J.D. and Anderson, A.B. (1983), *Handbook of Survey Research*, Quantitative Studies in Social Relations, Academic Press, New York, NY.

Simons, R. (2000), *Performance Measurement & Control Systems for Implementing Strategy: Text & Cases*, Prentice Hall, Upper Saddle River, NJ.

Ukko, J., Karhu, J. and Rantanen, H. (2007), "How to communicate measurement information successfully in small and medium-sized enterprises: a regression model", *International Journal of Information Quality*, Vol. 1 No. 1, pp. 41-59.

Voss, C., Tsikriktsis, N. and Frohlich, M. (2002), "Case research in operations management", *International Journal of Operations & Production Management*, Vol. 22 No. 2, pp. 195-219.

Yin, R.K. (2003), *Case Study Research: Design and Methods*, 3rd ed., Sage Publications, Thousand Oaks, CA.

Appendix

Item	Factor loading	Topic (category prior to analysis)
Instrument wording: the following statements relate to the Performance Measurement System and its measures. Please indicate your opinion on each statement on a scale from 1 to 7, where 1 = Completely disagree, 2 = Disagree, 3 = Slightly disagree, 4 = Neither disagree or agree, 5 = Slightly agree, 6 = Agree, 7 = Completely agree		
Proper System Design (α = 0.935)		
The quality of the measurement data is good (Franco and Bourne, 2003)	0.860	Metrics
Measures are understandable (Franco and Bourne, 2003; Melnyk *et al.*, 2004; de Waal, 2003; Bourne *et al.*, 2005)	0.779	Metrics
There is an appropriate amount of measurement targets (Melnyk *et al.*, 2004; Johnston and Clark, 2008)	0.766	Management
I can predict future outcomes with the measures (Franco and Bourne, 2003; Melnyk *et al.*, 2004; Radnor and Barnes, 2007)	0.759	Metrics
Measures reflect areas that are relevant to managing maintenance operations (Franco and Bourne, 2003)	0.743	Metrics
Measurement targets are set appropriately (Johnston and Clark, 2008; de Leeuw and van den Berg, 2011; Ittner and Larcker, 2003)	0.734	Management
I understand how the measurement data are generated (Franco and Bourne, 2003; de Waal, 2003)	0.701	Metrics
Relationships between measures are clear (Bourne *et al.*, 2005; Johnston and Clark, 2008; Neely and Bourne, 2000)	0.695	Metrics
There is an appropriate amount of measures (Franco and Bourne, 2003; Johnston and Clark, 2008; Kaplan and Norton, 1992)	0.662	Reporting
Measures relate to the strategic objectives of the company (Franco and Bourne, 2003; Johnston and Clark, 2008; de Leeuw and van den Berg, 2011; Ittner and Larcker, 2003)	0.615	Metrics
Training (α = 0.898)		
I have been trained to use the Performance Measurement System (Franco and Bourne, 2003; de Leeuw and van den Berg, 2011)	0.902	Training
What the measures mean has been explained to me (Franco and Bourne, 2003)	0.854	Training
I have instructions for the use of the Performance Measurement System[a]	0.824	Training
Reporting (α = 0.541)		
The Performance Measurement System report is easily accessible (Franco and Bourne, 2003; de Leeuw and van den Berg, 2011)	0.761	Reporting
The Performance Measurement System report is published frequently enough (Franco and Bourne, 2003; de Leeuw and van den Berg, 2011)	0.750	Reporting
(Deleted items)		
I prefer to use another measurement report instead of the Performance Measurement System (Reverse scale) (Bourne *et al.*, 2005)		Reporting

(*continued*)

Table AI. Summated scale (Cronbach's α)

Item	Factor loading	Topic (category prior to analysis)
Instrument wording: to what extent do you use the Performance Measurement System for the following purposes? Please indicate the extent of use for each purpose on a scale from 1 to 5, where 1 = Not at all, 2 = Little, 3 = Somewhat, 4 = Much, 5 = Very much		
Inspect and Improve (*α* = *0.961*)		
To follow trends in technician-level measures[a]	0.862	Control
To gain an overall picture of how well my entire team of technicians is performing[a]	0.801	Control
To identify development areas for technicians[a]	0.791	Education and Learning
To gain concrete information on how well each of my technicians is performing[a]	0.777	Control
To evaluate my entire team of technician's achievement of targets[a]	0.751	Control
To show technicians areas on which they should focus[a]	0.743	Signalling
To monitor targets set for my entire team of technicians[a]	0.739	Control
To coach technicians on how they could improve their work[a]	0.692	Education and Learning
Decision Making (*α* = *0.938*)		
To coordinate resources in my team of technicians[a]	0.884	Decision Making
To plan operations in my team of technicians[a]	0.873	Control
To make business decisions in my team of technicians[a]	0.811	Control
To allocate technician work in my team of technicians[a]	0.794	Decision Making
Motivate (*α* = *0.794*)		
To create a positive competitive setting between technicians[a]	0.817	Control
To motivate technicians[a]	0.643	Control
(Deleted items)		
To set targets for technicians[a]		Control
To evaluate technicians' achievement of targets[a]		Control
To give positive feedback to technicians[a]		Control
To follow trends in team-level measures[a]		Control
To communicate company goals to technicians[a]		Signalling
To learn the effects of my decisions[a]		Education and Learning
To improve my own management work[a]		Education and Learning
Instrument wording: the following statements relate to your team of technicians. Please indicate your opinion on each statement on a scale from 1 to 7, where 1 = Completely disagree, 2 = Disagree, 3 = Slightly disagree, 4 = Neither disagree or agree, 5 = Slightly agree, 6 = Agree, 7 = Completely agree		
Improve Competitiveness (*α* = *0.845*)		
In our team of technicians, we know how our company is planning to be competitive (Flynn and Flynn, 2004)	0.836	Communication
We have an overall picture of how well our team of technicians is performing[b]	0.758	Control
In our team of technicians, we believe that operational improvement is our responsibility (Douglas and Fredendall, 2004)	0.744	Improvement
Motivation (*α* = *0.822*)		
In our team of technicians, we like challenges (Jambulingam *et al.*, 2005)	0.827	Motivation

Table AI.

(*continued*)

Item	Factor loading	Topic (category prior to analysis)
In our team of technicians, we are very ambitious about our work (Jambulingam *et al.*, 2005)	0.814	Motivation
In our team of technicians, we consider ourselves to have high motivation towards work (Jambulingam *et al.*, 2005)	0.626	Motivation
Goal Communication (α = 0.821)		
Company goals, objectives and strategies are communicated to our team of technicians (Flynn and Flynn, 2004)	0.846	Communication
In our team of technicians, we understand the long-term competitive strategy of our company (Flynn and Flynn, 2004)	0.799	Communication
We know if our team of technicians meets its targets[b]	0.669	Control
(Deleted items)		
Strategies and goals are communicated primarily to my superiors (Reverse scale) (Flynn and Flynn, 2004)		Communication
Technicians in our team of technicians are a group of hard-working individuals (Jambulingam *et al.*, 2005)		Motivation
We are able to influence the way our team of technicians operates[b]		Control
In our team of technicians, we continuously improve our operations (Douglas and Fredendall, 2004)		Improvement
In our team of technicians, we analyse our operations to look for ways of doing a better job (Douglas and Fredendall, 2004)		Improvement
Instrument wording: Please evaluate performance of your team of technicians on following aspects on a scale from 1 to 7, where 1 = Very poor, 2 = Poor, 3 = Somewhat poor, 4 = Average, 5 = Somewhat good, 6 = Good, 7 = Very good		
Efficiency (α = 0.824)		
Productivity (Dean and Snell, 1991)	0.846	Performance
Accuracy of maintenance (Dean and Snell, 1991)	0.752	Performance
Completion time of maintenance tasks (Dean and Snell, 1991)	0.736	Performance
Utilisation of technician resources (Dean and Snell, 1991)	0.658	Performance
Dedication (α = 0.699)		
Team motivation (Dean and Snell, 1991)	0.892	Performance
Maintenance work quality (Dean and Snell, 1991)	0.746	Performance
Extra work (α not applicable – single item)		
Extra work by technicians[a]	0.922	Performance

Notes: [a]Meetings and interviews. [b]Authors

Table AI.

Corresponding author
Professor Jan Holmström can be contacted at: jan.holmstrom@aalto.fi