Performance measurement and management: theory and practice

Guest Editors: Mike Bourne, Steven Melnyk and Umit S. Bititci

2009 Editorial board
2010 Guest editorial
2022 The relationship regulator: a buyer-supplier collaborative performance measurement system
Vieri Maestrini, Veronica Martinez Andy Neely, Davide Luzzini, Federico Caniato and Paolo Maccarrone

2040 The impact of supplier performance measurement systems on supplier performance: a dyadic lifecycle perspective
Vieri Maestrini, Davide Luzzini, Federico Caniato, Paolo Maccarrone and Stefano Ronchi

2062 Missing performance management and measurement aspects in performance-based contracting: a systematic process-based literature analysis of an astonishing research gap
Andreas H. Glaß, Florian U. Henne and Michael Essig

2096 Operations flow effectiveness: a systems approach to measuring flow performance
Mohamed Afy-Shararah and Nicholas Rich

2124 The performance effects of complementary management control mechanisms
Henrik Nielsen, Thomas Borup Kristensen and Lawrence P. Grasso

2149 Controlling lean manufacturing in multidivisional organisations: highlighting local interests and constraints
Martijn Pieter van der Steen and Sandra Tillema

2169 The use of management control and performance measurement systems in SMEs: a levers of control perspective
Biljana Pešalj, Andrey Pavlov and Pietro Michel

2192 Predicting performance – a dynamic capability view
Marcus F. Hasegan, Sai Sudhakar Nudurupati and Stephen J. Childe

2214 A decision theory perspective on complexity in performance measurement and management
Anthony Alexander, Maneesh Kumar and Helen Walker

ISBN 978-1-78973-199-6

www.emeraldinsight.com/loi/ijopm
ASSOCIATE EDITORS

Par Ahlstrom
Stockholm School of Economics, Sweden
Arash Azadegan
Rutgers University, USA
Umit Sezer Bititci
Heriot Watt University, UK
Christoph Bode
University of Mannheim, Germany
Steve Brown
Southampton Business School, UK
Raffaella Cagliano
Politecnico di Milano, Italy
Simon Croom
University of San Diego, USA
Giovani da Silveira
Haskayne School of Business, University of Calgary, Canada
Joy Field
University of Warwick, UK
Kai Foerstl
German Graduate School of Management & Law, Germany
Jose Arturo Garza-Reyes
University of Derby, UK
Janet Godsell
Warwick Business School, UK
Jan Hartley
Bowling Green State University, USA
Alex Hill
Kingston University, UK
Canan Kocabasoglu Hillmer
City University London, UK
Mathias Holweg
University of Oxford, UK
Juliana Hsuan
Copenhagen Business School, Denmark
Jayanth Jayaram
Darla Moore School of Business, University of South Carolina, USA
Vaidyanathan Jayaraman
Great Lakes Institute of Management, India
Jeff Jia
University of Bristol, UK
David Johnston
York University, Canada
Patrik Jonsson
Chalmers University of Technology, Sweden
Benn Lawson
Judge Business School, University of Cambridge, UK
Kevin Linderman
University of Minnesota, USA
Steven Melnyk
Michigan State University, USA
Jason Miller
Michigan State University, USA
Sachin Modi
Wayne State University, USA

Adegoke Oke
Arizona State University, USA
Jan Olhager
Lund University, USA
Thanos Papadopoulos
The University of Kent, UK
Antony Paullraj
University of Nottingham Ningbo, People’s Republic of China
Damien James Power
The University of Melbourne, Australia
Nick Rich
Swansea University, UK
Jens Roehrich
University of Bath, UK
Ken Schultz
Air Force Institute of Technology, USA
Stefan Seuring
University of Kassel, Germany
Palie Smart
University of Bristol, UK
Rui Sousa
Universidade Catolica Portuguesa, Portugal
Martin Spring
University of Lancaster, UK
Mark Stevenson
University of Lancaster, UK
Stephan Vachon
Western University, Canada
Dirk Pieter van Donk
University of Groningen, The Netherlands
Ann Vereecke
Vlerick Business School, Belgium
Frank Wiengarten
Esade Business School, Universitat Ramon Llull, Spain
Christina Wong
Hong Kong Polytechnic, Hong Kong
Qinghua Zhu
Shanghai Jiao Tong University, People’s Republic of China

AD-HOC ASSOCIATE EDITORS

Tim Baines
Aston University, UK
Paul D. Cousins
University of Bristol, UK
Andreas Größler
University of Stuttgart, Germany
Robert Handfield
North Carolina State University, USA
Jose A.D. Machuca
University of Sevilla, Spain
Maria J. Montes-Sancho
Universidad III de Madrid, Spain
Zoe Radnor
Leicester University, UK
Amrik Sohal
Monash University, Australia

The full Editorial Advisory Board can be found listed on the IJOPM website:
www.emeraldgrouppublishing.com/ijopm.htm
Performance measurement and management: theory and practice

Introduction

Performance measurement has been a key theme of this journal for many years (Neely et al., 1995; Neely, 2005). This should not come as a surprise since performance measurement and management plays a critical role in the operation of any organisation, be it a factory, business, hospital or school. As noted by Magretta and Stone (2002), performance measures are critical because they enhance communication — they enable the organisation to address the following critical question – “Given our mission, how is our performance going to be defined?” Yet, it should also come as a surprise that in spite of this longevity of focus, this topic is still surprised by a great deal of confusion and conflict. For example, until more recently we have not agreed on what we mean by performance measurement (Franco-Santos et al., 2007; Bourne and Bourne, 2011; Melnyk et al., 2004, 2014) and performance management (Bourne and Bourne, 2011; Melnyk et al., 2004, 2014). This confusion is present in both academic research and in the practitioner press. Confounding this confusion is the tight interrelationship that exists between theory and practice.

It has been long argued that many of the developments in performance measurement have come from practice (Johnson, 1972, 1975, 1978, 1981; Wilcox and Bourne, 2003) and this practice has informed the academic performance measurement literature. A very clear example of this phenomenon is the Balanced Scorecard, originally developed at Analogue Devices (Schneiderman, 2001) and subsequently made widely accessible by Kaplan and Norton (1992). But although this type of research is to be applauded, it is not enough. Ultimately, it suffers from being a theoretical, using Sutton and Staw (1995) categorisation scheme. Such research, while important for improving performance and the ability of the firm to record and monitor activities, does not build causality or help to explain or, as is becoming more important in today’s dynamic environment (Nudurupait et al., 2016), predict and deal with the increasing complexity now being faced by researchers and managers alike.

We need to build theory, whether from practice or other areas of business research, and validate our theory through empirical research to develop a deeper understanding and platform for the future development of the field. Furthermore, we need to build theory that operates at the various levels of the organisation – inter-organisational (relevant for a world where supply chain management is becoming so important); organisational at macro or top management levels (the focus of much of the current body of research in this area) and the lower levels of the organisation (i.e. at the function, group and, ultimately, the individual levels – the micro levels). We need to build theories that are anchored not only in the economic world but that also recognise the increasing importance of behavioural issues. So, for this special issue, we were particularly interested in papers that explore the interface between theory and practice and that add richness, as described in the preceding discussion, to the resulting theoretical developments.

But performance measurement and management are broad subjects and performance measurement and management systems (PMMS) have a wide scope. They include the top down processes of aligning the business with strategic direction as promulgated by process approaches to the development and implementation of PMMS (e.g. Kaplan and Norton, 1993; Neely et al., 1996; Bititci et al., 1997; Neely et al., 2000) and to the bottom up use of performance measurement in lean (Bhasin and Burcher, 2006; Krafčík, 1988). They include the dynamics of people and teams, interactions between department, relationships between
parent and subsidiary organisations as well as performance measurement and management in the supply chain.

Performance measurement and management is not only the preserve of the private sector. Articles in this journal have reported on the impact of PMMS in both public and third sector organisations (Greatbanks and Tapp, 2007; Moxham and Boaden, 2007). This special issue too reflects this pervasiveness of PMS, despite the fact that the papers published here happen to have focused on the private sector. We have a number of different perspectives on performance measurement and management. This includes three papers on buyer–supplier relationships, one on collaborative performance measurement systems, one the impact of supplier performance management systems (PMS) and one on contracting. We have five papers looking at the complexities and interactions between different elements of the performance measurement system inside organisations with two of these focusing on lean. The final paper develops a framework for swift and even flow.

So in the rest of this editorial we will discuss the following. Let us begin with an overview of the early literature and the development of the four phases of PMMS before turning our attention to the need for theory to focus on the use of PMMS in directing and managing organisations. Then we will move on to the range of theory used in empirical research and in the papers presented here before going on to describe the contributions made in this special issue. We will then discuss the contribution of the Ferriera and Otley (2009) framework before suggesting future direction for theory development.

The use of theory in performance measurement and management

Early literature on PMMS focused on the four phases of PMMS, design, implementation, use and refresh (Neely et al., 2000; Bourne et al., 2000). There is a strong argument that the debate has now moved on from the design and implementation of PMMS to its use, Franco-Santos and Bourne, 2005), impact (Franco-Santos et al., 2012) and emergence (Pavlov and Bourne, 2011).

The work on approaches (Dixon et al., 1990) and processes for the design and implementation of PMMS (Neely et al., 1997; Bititci et al., 1997; Olve et al., 1999) initially adopted the three tests on feasibility, usability and utility (Platts, 1993) before considering the success and failure of the design to implementation phase (Bourne et al., 2002; Bourne, 2005) and the quality of implementation. One of the overarching pieces here was the Bititci et al. (2006) paper which identified the need for different cultures and management styles when moving from the design and implementation phases to measurement in use phases. We would argue that it is possible to theorise about the design and implementation phases but in reality, these are as much about the implementation of change (Bourne et al., 2004) as anything else. There have also been interesting insights into the formal approaches to refreshing measurement systems (Kennerley and Neely, 2003) and more recently addressing the issue of how to keep measurement systems up to date (Melnyk et al., 2014), but we strongly suggest that there is now a compelling need to develop theory around the continuing use and emergent development of PMMS.

In their paper, Franco-Santos et al. (2012) identified that the empirical work undertaken to research the impact of performance measurement systems on performance had focused on the use of six groups of theories, although it should be noted that a third of the empirical papers reviewed in their paper used no theoretical basis at all. The six groups of theories were:

(1) agency theory (Eisenhardt, 1989; Feltham and Xie, 1994; Jensen and Meckling, 1976; Jensen and Murphy, 1990);
(2) contingency theory (Donaldson, 2001; Hayes, 1977; Otley, 1980);
(3) resource-based view of the firm (Barney, 1991; Day, 1994);
cognitive and information processing theories (Miller, 1956; Simon, 1976; Talyer, 2010; Kunda, 1990);

goal setting theory (Locke and Latham, 1990); and


In this special issue, we have papers too that use theories matching three of the six groups of theories identified in Franco-Santos et al. (2012). Papers published here are using resource-based view (2), cognitive theories and decision making (2) and contingency theory. But we also have papers focusing on theories more relevant to PMMS in operations (performance-based contracting (PBC) and swift even flow) as well as theories related to complementarity in PMMSs, and extending the concept of levers of control (Simons 1991). The papers in this special issue are summarised in Table I. We will briefly describe these papers next.

The first paper by Vieri Maestrini, Veronica Martinez, Andy Neely, Davide Luzzini, Federico Caniato and Paolo Maccarrone looks at how buyers and suppliers can collaborate in their use of their performance measurement system. In this paper, they present a tool they have called the “Relationship Regulator” which they develop, test and refine. The development was based on the literature and empirical research, whilst the testing was undertaken through workshops and feedback obtained from semi-structured interviews. This is theory building and testing of a collaborative approach to performance measurement.

The second paper by Vieri Maestrini, Davide Luzzini, Federico Caniato, Paolo Maccarrone and Stefano Ronchi is an impact paper in that it researches the impact of supplier performance measurement systems on supplier performance through hypothesis testing using survey data. However, the interest in this paper is the use of resource orchestration theory (ROT) as the theoretical framework for their analysis. They found ROT to be a suitable theoretical framework to explain the role of a mature supplier performance measurement system in orchestrating the suppliers.

The third paper by Andreas Glas, Florian Henne and Michael Essig is a literature review on the intersection of PBC and performance measurement and management. The review highlights the performance measurement and management gap in PBC identifying for research opportunities: strategic alignment (which the authors consider astonishing as PBC is supposed to be outcome orientated), action and refreshing, performance monitoring and reporting, and other aspects including the selection of appropriate KPIs.

The fourth paper by Mohamed Afy-Shararah and Nicholas Rich creates a model that captures the design elements of high performance operating management systems for swift even flow. The work builds on ten longitudinal case studies selected from the UK’s high value manufacturing sector highlighting the links between policy deployment, operational control and improvement to open collaborative partnerships.

The fifth paper by Henrik Nielsen, Thomas Kristensen and Lawrence Grassol is the first of two papers focusing on lean. The paper uses survey research to investigate the impact of social control mechanisms, behavioural control mechanisms and their complementarity on firm performance using data collected from over 4,000 subjects in nearly 700 facilities. The respondents to the survey behind this work were identified from through the “Shingo Prize”.

The sixth paper by Martijn van der Steen and Sandra Tillema takes a different approach and looks at lean implementation. Using case studies in three subsidiary companies, they suggest that lean can be severely constrained by the parent organisations accounting based control systems. The paper demonstrates how external context creates local conditions that may be detrimental to lean in manufacturing.

The seventh paper by Bijnan Pesalji, Andrey Pavlov and Pietro Micheli uses the levers of control framework from Simons (1991) to investigate practices in a Dutch-based SME. The paper advances our understanding of the use of technical and social aspects of
performance management and suggests that performance management requires the active and continuous use of all four control mechanisms that comprise the levers of control approach.

The eighth paper by Marcus Hasegan, Sai Nudurupati and Stephen Childe reports on the use of action research for developing dynamic performance measurement systems with real-time controls on the production lines to study the impact. The paper explains how the use of tacit knowledge and modelling were used in developing effective cause and effect analysis.

The last paper by Anthony Alexander, Maneesh Kumar and Helen Walker considers the application of decision theory under volatility, uncertainty, complexity and ambiguity (VUCA).

<table>
<thead>
<tr>
<th>Authors</th>
<th>Subject</th>
<th>Theoretical lens</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vieri Maestrini, Veronica Martinez, Andy Neely, Davide Luzzini, Federico Caniato and Paolo Maccarrone</td>
<td>The development of a smart tool to allow buyer–supplier dyads to collaborate on relationship performance</td>
<td>Framework building and testing</td>
<td>A tested and improved buyer–supplier tool</td>
</tr>
<tr>
<td>Vieri Maestrini, Davide Luzzini, Federico Caniato, Paolo Maccarrone and Stefano Ronchi</td>
<td>The impact of supplier performance measurement systems on supplier performance</td>
<td>Resource orchestration theory (Hitt et al., 2011)</td>
<td>Resource orchestration theory explains the role of a mature supplier performance measurement system in orchestrating suppliers</td>
</tr>
<tr>
<td>Mohamed Afy-Shararah and Nicholas Rich</td>
<td>A systems approach to measuring and managing flow performance in high value manufacturing</td>
<td>Systems theory</td>
<td>Design elements of a high performing operational management system for swift even flow</td>
</tr>
<tr>
<td>Henrik Nielsen, Thomas Kristensen and Lawrence Grasso</td>
<td>The performance effects of complimentary control mechanisms in the implementation of lean</td>
<td>Complementarity of management control systems (MCS)</td>
<td>Deeper understanding of how the complementary effects of lean manufacturing impact on firm performance</td>
</tr>
<tr>
<td>Martijn van der Steen and Sandra Tillema</td>
<td>Implication of context on lean implementation</td>
<td>Contingency</td>
<td>Better understanding of how organisational controls, especially accounting control, impact on lean</td>
</tr>
<tr>
<td>Biljana Pesalj, Andrey Pavlov and Pietro Micheli</td>
<td>A levers of control perspective on the use of MCS and performance measures in SMEs</td>
<td>Levers of control (Simons, 1991)</td>
<td>Extending the concept of levers of control by identifying practices and the importance of maintaining balance</td>
</tr>
<tr>
<td>Marcus Hasegan, Sai Nudurupati and Stephen Childe</td>
<td>A dynamic performance measurement system (DPMS)</td>
<td>Dynamic capabilities</td>
<td>An understanding of using cause and effect relationships in manufacturing to better predict performance</td>
</tr>
<tr>
<td>Anthony Alexander, Maneesh Kumar and Helen Walker</td>
<td>Decision theory and the emergence with complexity in performance measurement and management</td>
<td>Decision theory</td>
<td>Dealing with complexity and emergence, enabling managers in positions of authority to make decisions</td>
</tr>
</tbody>
</table>
Drawing on the Cynefin framework (Snowden, 2000, 2002; Snowden and Boone, 2007), this paper develops the performance alignment matrix (Melnyk et al., 2014) drawing on interview research in seven case studies. The approach is designed to deal with complexity and emergence enabling managers in the positions of authority to take decisions. This paper highlights the need for organisations to adjust their performance measurement and management system over time to adapt to the external environment as a way of reformulating strategy, promoting intended behaviour and organisational learning.

So, from a theory and practice perspective, we must conclude that the special issue has attracted a wide variety of papers, research methods and applications from different contexts. The papers contribute to our understanding of performance measurement and management in buyer–supplier relationships, performance contracting and operational flow. The papers create insight into the issue of complimentary or conflicting aspects of control systems, especially in a lean setting and we have insights into balancing controls in SMEs as well as how to deal with VUCA environments. In the next section, we will focus on discussion on theory in PMMS.

**Discussion**

If we review the theoretical approach taken both the papers in this special issue and by those empirical pieces identified in the Franco-Santos et al. (2012) literature review, it must be noted that they are not theories of performance measurement and management, but more general management theories applied to this subject area. What we are currently lacking is an underpinning theory to help us advance the field.

However, the field is not completely without a theoretical underpinning, the Ferriera and Otley (2009) theoretical framework is a very useful step in this direction as it gives us a framework and a set of 12 questions to help us analyse PMMSs. The 12 questions developed by Ferriera and Otley (2009, pp. 266-267) are as follows:

1. What is the vision and mission of the organisation and how is this brought to the attention of managers and employees? What mechanisms, processes and networks are used to convey the organisation’s overarching purposes and objectives to its members?
2. What are the key factors that are believed to be central to the organisation’s overall future success and how are they brought to the attention of managers and employees?
3. What is the organisation structure and what impact does it have on the design and use of PMSs? How does it influence and how is it influenced by the strategic management process?
4. What strategies and plans has the organisation adopted and what are the processes and activities that it has decided will be required for it to ensure its success? How are strategies and plans adapted, generated and communicated to managers and employees?
5. What are the organisation’s key performance measures deriving from its objectives, key success factors and strategies and plans? How are these are specified and communicated and what role do they play in performance evaluation? Are there significant omissions?
6. What the level of performance does the organisation need to achieve for each of its key performance measures (identified in the above question), how does it go about setting appropriate performance targets for them and how challenging are those performance targets?
7. What processes, if any, does the organisation follow for evaluating individual, group and organisational performance? Are performance evaluations primarily objective,
subjective or mixed and how important are formal and informal information and controls in these processes?

(8) What rewards – financial and/or non-financial – will managers and other employees gain by achieving performance targets or other assessed aspects of performance (or, conversely, what penalties will they suffer by failing to achieve them)?

(9) What specific information flows – feedback and feedforward – systems and networks has the organisation in place to support the operation of its PMSs?

(10) What type of use is made of information and of the various control mechanisms in place? Can these uses be characterised in terms of various typologies in the literature? How do controls and their uses differ at different hierarchical levels?

(11) How have the PMSs altered in the light of the change dynamics of the organisation and its environment? Have the changes in PMS design or use been made in a proactive or reactive manner?

(12) How strong and coherent are the links between the components of PMSs and the ways in which they are used.

We should reflect that the framework, whilst informed by the management control literature, was developed inductively and informed by the experience of case study research (Ferriera and Otley, 2009, p. 276), which is why it has such a strong link to practice and why it is so useful as a holistic tool for examining the structure, operation and use of PMSs in organisations.

In many ways, this framework follows some of the early research works (Neely et al., 1996) and publications (Kaplan and Norton, 1993, 2001) on PMMS, whereby the vision and mission of the organisation is translated into a strategy that is operationalised through the performance measurement and management systems. However, the Ferriera and Otley (2009) framework progresses far beyond this as it includes the development of targets, the processes of evaluation, information flow and link to rewards. Further, question 11 alludes to the dynamic nature of PMMSs and recognises that they evolve over time; so their development is not solely determined through interventions such as redesign initiatives.

Although this is an extremely useful theoretical framework, it is not a theory of performance measurement or performance management. To develop such a theory, we need to understand the mechanisms at play in the organisation when performance measures are being used to manage activities, changes of activity and future direction of the organisation. There have been calls for such a meta theory (Franco-Santos et al., 2012; Bittitci et al., 2018) and we will suggest in the next section one possible direction to take in developing such an approach.

Towards one theory of performance measurement and management

The Ferriera and Otley (2009) framework described above assumes that PMMSs are systems. In reality, they may be “systems of systems” (Bourne et al., 2018) but this framework does align with a recent call to take a more systemic approach to developing theory in performance measurement and management (Bittitci et al., 2018). These systems operate through practices and routines in organisations and it is to this subject we turn next.

If we reflect on the current theories used to inform PMMS research, it can be argued that they ignore the mechanisms by which the PMMS operates. From Franco-Santos et al.’s (2012) six theories, four of them (agency, cognitive, goal setting and equity theories) focus on the influence of system parameters on what is usually considered to be the individual decision maker. The other two focus on the impact of the wider environment, be
this the external (contingency theory) or internal (resource-based view) context. What we are not arguing here that these are not useful lenses to take in analysing the PMMSs in organisations, we are simply suggesting that adopting an alternative approach that takes a systems perspective by focusing of the operating mechanisms of PMMS could give use new insights.

One such approach is the adoption of routines (Pavlov and Bourne, 2011). In their paper, the authors argue that performance measurement triggers, guides and intensifies the search for solutions to improve the performance of the organisation (see Figure 1). As this happens in performance review meetings (regular events where groups of managers meet to review, evaluate and act on performance information (Martinez et al., 2010); Pavlov and Bourne, 2011) when routines are developed which evolve into mechanisms used in the management of the organisation. This approach takes us away from the concept of the individual decision maker to the domain of multiple decision makers (albeit in a situation where some have considerably more power and influence than others). Understanding the mechanisms at play here would give us different insights.

In reality, organisations have multiple situations where performance is reviewed. These may be formal board meetings, operational planning meetings, sales management meetings, project management meetings, etc. Decisions may be made by individual decision makers acting on their own (although we would suggest that this happens far more infrequently than one would surmise from the focus of research in the management literature) but implementing action invariably requires involving others. This suggests (in all but the smallest organisations) a series of links between individuals at different levels of the organisation. Individual managers involved in running one routine (performance review meeting) will usually be participants in a higher level performance review meeting, whilst the people attending this manager’s meeting may well be running their own routine (performance review meeting) at a lower level in the organisation. In this situation, routines being influenced and guided by other routines (see Figure 2) together with the individual managers’ membership of performance review meetings being the conduit of the PMMS between different levels in the organisation (see Figure 3).

Because these mechanisms are routines, they develop over time as the mental models of what is happening influences the pattern of behaviours and the pattern of behaviours influence the mental models. This is not only sense making in practice, but probably also how emergent strategy develops and is implemented outside the formal mechanisms for strategy development in most organisation.

**Figure 1.**
Effects of PM on organisational routines

*Source: Pavlov and Bourne (2011)*
Figure 2. Effects of the structure of PMMS on organisational routines

Figure 3. The overlap of points of membership in the hierarchy of organisation routines
Conclusion

The domain of performance measurement and management continues to develop and, if performance measurement and management evolve to enable us to control new and emerging forms of organisations in new and emerging contexts, then it will always continue to do so. In this special issue, we have seen the development of interesting new frameworks and understanding based on using theory to reflect on practice and practice to inform our development of theory.

However, we still believe that future research in this field would benefit from adopting a more systems-based approach to understanding the mechanisms at play in PMMS. We have suggested that understanding these mechanisms in practice would be beneficial to the field and would give us a platform for future research. But, if and when we develop this better understanding, we suspect that we will still be using other more encompassing management theories, such as those used and identified here, to understand the pressures and influences on the mechanism.

Mike Bourne
Centre for Business Performance, Cranfield School of Management, Cranfield University, Bedford, UK

Steven Melnyk
Eli Broad College of Business, Michigan State University, East Lansing, Michigan, USA, and

Umit S. Bititci
Department of Business Management, Heriot-Watt University, Edinburgh, UK

References


Further reading


The relationship regulator: a buyer-supplier collaborative performance measurement system

Vieri Maestrini
School of Management, Politecnico di Milano, Milan, Italy
Veronica Martinez
Institute for Manufacturing, University of Cambridge, Cambridge, UK
Andy Neely
Cambridge University, Cambridge, UK
Davide Luzzini
Department of Marketing, Operations & Supply, Escuela de Alta Direccion y Administracion, Barcelona, Spain
Federico Caniato
School of Management, Politecnico di Milano, Milan, Italy, and
Paolo Maccarrone
Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy

Abstract
Purpose – The purpose of this paper is to propose an innovative buyer-supplier performance measurement system (PMS) (called relationship regulator – RelReg), aimed at stimulating collaboration on mutual performance. The RelReg is described throughout the phases of its lifecycle: first, design features and visual representation of the new measurement framework are reported; second, guidelines on how to implement, use and review the system are provided, highlighting the role of the buyer and the supplier at each step.
Design/methodology/approach – A theory building and testing approach is applied. The RelReg developed features primarily ground on previous scientific contributions matched with empirical evidence collected through case studies, workshops and focus groups. The resulting conceptual model is then validated through a dyadic buyer-supplier case study.
Findings – Two conceptual frameworks are provided: the RelReg dashboard – a multidimensional PMS; and the RelReg lifecycle – set of activities to be performed by both the buyer and the supplier all along the adoption process. Moreover, empirical insights on relevant issues to be considered when adopting the RelReg are reported.
Originality/value – The RelReg represents an innovative and smart tool, allowing buyer-supplier dyads to collaborate on relationship performance.

Keywords Performance measurement, Buyer-supplier relationships

Paper type Conceptual paper

1. Introduction
Over the years, there has been a generalized tendency to increase management vision and control, with companies seeking to control over inter-firm processes and relationships. Several authors have therefore suggested that traditional intra-organizational performance measurement systems (PMs) need to be broadened, with the development of external supply chain PMs (SCPMs), crossing company boundaries (Gunasekaran et al., 2004; Chae, 2009; Gunasekaran and Kobu, 2007). Easier said than done.

Three factors need to be considered in nowadays business environment. First, supply chains (SCs) are becoming more and more fuzzy: rather than being mutually exclusive
chains, they appear as interconnected and overlapping networks, where companies are immersed and linked through diverse types of relationship (Lambert and Pohlen, 2001; Rice and Hoppe, 2001). Focus and choice are essential when extending the measurement process beyond company boundaries, yet often complex. Second, organizational skills are critical to design and take full advantage of an SCPMS. Although purchasing, SC and customer service functions have increased their managerial capabilities in recent times (Luzzini and Ronchi, 2016), they still rarely display and follow formal strategies (Hesping and Schiele, 2015). Third, a reliable and robust information system infrastructure is critical for a successful implementation of external SCPMSs (Nudurupati et al., 2011). This requires technological knowledge, resources and investments in order to tailor the ICT systems to the company specificities.

In the last 15 years, internal PMS literature has progressively moved from the measurement system design to its implementation (Bourne et al., 2000, 2002), use (Henri, 2006; Koufteros et al., 2014) and review (Braz et al., 2011). External SCPMS literature has not experienced this evolution yet. Contributions are still strongly focused on the “what to measure issue,” with a profusion of studies on which performance dimension to tackle (Kannan and Tan, 2002; Gunasekaran et al., 2004) and how to select relevant metrics (Cai et al., 2009; Agarwal et al., 2006). Besides, in most cases, only the point of view of the buyer company evaluating its suppliers is considered, thus neglecting two elements. First, suppliers do measure performance of their customers by means of customers PMSs in many cases. Second, actively considering the perspectives of both parties is critical to evaluate the effectiveness of the measurement tool. Finally, suppliers PMSs are generally viewed as diagnostic tools for monitoring, which the buyer puts in place to control its supply base with an evaluation purpose (Henri, 2006). The role of the measurement system in enabling mutual collaboration on performance has not been thoroughly investigated so far (Koufteros et al., 2014; Melnyk et al., 2014).

The present study aims at challenging limitations of extant literature by building and testing an innovative framework allowing the buyer-supplier collaboration on mutual performance. We call it the relationship regulator (RelReg). The RelReg is explained all along its lifecycle elements (i.e. design, implementation, use and review phases), highlighting the role of both parties at each step.

The structure of this paper is organized as follows: Section 2 reports a review of extant scientific literature on the subject, addressing the streams of SCPMS, supplier evaluation and the buyer-supplier relationship evaluation. Section 3 resumes the goal of the paper and the methodology adopted. In Section 4, RelReg is described in its constituent elements, highlighting both the design features and guidelines to follow along the implementation, use and review. Besides, empirical evidence from a first buyer-supplier dyadic case study is discussed. Section 5 reports a critical discussion of the pros and cons of RelReg as emerging from the case study. Conclusions end the paper.

2. Literature review

Starting from the late 1990s (Van Hoek, 1998; Beamon, 1999), several authors in the academic literature have reported on studies of the development of PMSs addressing the evaluation of activities outside legal company boundaries. Hald and Ellegaard (2011) identify three converging and overlapping streams of research, according to the scope of the system they address and the labels used: SCPMS tackling SC processes and practices (Gunasekaran et al., 2001, 2004; Angerhofer and Angelides, 2006); supplier evaluation focusing on first tier suppliers (Simpson et al., 2002; Kannan and Tan, 2002; Luzzini et al., 2014); the buyer-supplier relationship assessment, focusing on soft aspects like mutual commitment, integration, trust, etc. (Giannakis, 2007; Ramanathan et al., 2011). For the sake of clarity, it is worth providing precise definitions of recurrent labels in this paper. Influenced by Neely et al’s (1995)
definition of PMS, we refer to external SCPMS as a set of metrics used to quantify the efficiency and effectiveness of inter-firm processes and relationships. From the perspective of a business-to-business company, we can eventually distinguish between suppliers PMSs (set of metrics used to quantify the efficiency and effectiveness of suppliers' actions) and customers PMSs (set of metrics used to quantify both the efficiency and effectiveness of customers' actions).

Within the broad area of external SCPMS, most studies address the pattern of the evaluating buyer company, adopting supplier PMSs to control and orchestrate its supply base. This implicitly uncovers two main limitations: the lack of insights on customer PMSs put in place by supplier companies to monitor their buyers performance: apart from a few comprehensive SCPMS tackling also downstream processes and relationships (e.g. Gunasekaran and Kobu, 2007; Bullinger et al., 2010), customers PMS are largely neglected, yet often used by companies’ customer service functions; and the paucity of contributions reporting also the point of view of the evaluated company: in assessing the effectiveness of the measurement process, it seems logical to take into account both the evaluating and evaluated company perspectives. On this behalf, it is interesting to note that the few studies jointly reporting the dyadic perspective actually highlight a strong dichotomy of perceptions between the two parties. Purdy et al. (1994) and Purdy and Safayeni (2000) report three main conclusions: the majority of suppliers feel that their effectiveness is not accurately reflected in the evaluation, which seems more a test of how much their companies look like the buyer; the evaluating buyer company did not utilize the information gathered through the audit process properly because in the end, their decisions were based only on price savings; and suppliers believe that the score reported is driven by bargaining power rules and does not result from a formal and objective evaluation process. Hald and Ellegaard (2011), by means of two longitudinal case studies, investigate how performance measurement information, traveling between the evaluating buyer and the evaluated suppliers, is shaped and reshaped in the evaluation process. The authors highlight that a harsh dialectic often arises between the two parties on the supplier PMS put in place.

Another characteristic of the extant scientific literature on external SCPMSs is the primary focus on the design process. Various models have been proposed over the years, like the SC balanced scorecard (Brewer and Speh, 2000; Bhagwat and Sharma, 2007); the SCOR framework (Sellitto et al., 2015; Gunasekaran et al., 2004; Arzu Akyuz and Erman Erkan, 2010); process-based approach (Chan and Qi., 2003); and suppliers’ scorecard (Kannan and Tan, 2002). In parallel algorithms and methodologies for metrics, selection and prioritization have been proposed like the analytic hierarchy process (AHP) (Cai et al., 2009; Cho et al., 2012), fuzzy AHP (Hong and Zhong-Hua, 2013), analytic network process (ANP) (Agarwal et al., 2006). On the other hand, empirical investigation on the effectiveness of previous frameworks and the analysis of other phases within their lifecycle (i.e. the implementation, use and review) are lacking. Nonetheless, several contributions on internal PMSs acknowledge that the implementation (Bourne et al., 2000, 2003), use (Henri, 2006; Koufteros et al., 2014) and review (Lohman et al., 2004; Braz et al., 2011) of a PMS are crucial determinants of its success, as important as a proper design.

Finally, previous contributions on SPMS have addressed the relationship between the system adoption and diverse relationship capabilities constructs, such as socialization mechanisms (Cousins et al., 2008), cooperation (Mahama, 2006), supplier’s integration (Carr and Pearson, 1999) and supplier’s commitment (Prahinski and Fan, 2007; Prahinski and Benton, 2004). However, these papers look at the role played by these relationship capabilities within the SPMS adoption - performance improvement path. This reflects the fact that traditional supplier PMS entails a merely evaluation purpose, with relationship capabilities (like collaboration), which may emerge as collateral factors. This paper tackles this issue by proposing an innovative buyer-supplier PMS, which intrinsically grounds on mutual collaboration to enhance performance.
3. Research aim and methodology
Reviewing the extant scientific literature, several signals suggest that the traditional pattern entailing the buyer company evaluating its suppliers appears as a highly constraining and limiting scheme. First, the fact that suppliers do actively measure some performance of their customers (e.g. forecasting accuracy, payment timeliness) is not taken into account. Second, it seems to prevent the development of relational capabilities (such as mutual commitment, social capital), which strongly affects the effectiveness of the measurement process. Within this paper, we challenge this unidirectional and diagnostic paradigm, by proposing the RelReg, an innovative framework aimed at stimulating a collaborative buyer-supplier performance measurement and management.

The RelReg entails a dyadic joint measurement of balanced performance dimensions: some addressing the supplier, some addressing the buyer and some others addressing both parties. The logic behind is to use the measurement tool to enable collaboration and continuous improvement on relationship performance. Acknowledging the critical role of all the phases within the SCPMS lifecycle, after presenting the RelReg design features, we highlight the key activities to be performed when implementing, using and reviewing the framework (Gutierrez et al., 2015; Bititci et al., 2006).

3.1 Theory building phase
The research is based on a two-step methodology (see Figure 1). The first step grounds on a theory building process, resulting into the development of the RelReg dashboard and lifecycle. The second step is based on a buyer-supplier dyadic case study and it is aimed at refining and providing a first attempt of the RelReg validation by synthetizing relevant issues to be addressed when adopting the tool.

The RelReg conceptualization (first step of the methodology) grounds on two pillars: the review of the extant scientific literature on intra- and inter-company performance measurement; and empirical data coming from the first author’s experience on scientific and applied research projects on supplier performance measurement and management, matured in the last five years. Directly inspiring the present research, the following can be mentioned: 30 case studies addressing buyer companies (manufacturing, retail, service sectors)
as measuring party, for a total of 88 interviews to managers in the purchasing, SC or logistics department; 22 case studies addressing supplier companies (manufacturing, retail, service sectors) as measured party, for a total of 61 interviews; participation to 9 academic conferences including tracks on SC/supplier performance measurement and management; participation as guest speaker to 3 practitioner conferences in the area of purchasing and supply management, holding speeches on supplier performance measurement and management; organization of one practitioners’ workshop on the topic; and design and execution of a buyer-supplier survey on the design, implementation, use and review of supplier PMSs, achieving a final sample of 147 dyadic responses. As a final ingredient, the co-authors’ long research experience on purchasing and supply management and on performance measurement and management has been important to develop and refine the framework.

3.2 Theory testing phase
The output of the previous theory building phase is the development of the RelReg dashboard (cf. Figure 2), and of the RelReg lifecycle (cf. Figure 4). Afterwards, a first dyadic buyer supplier case study has been performed, aimed at a preliminary validation of the RelReg and at gaining insights on its applicability. The following table reports relevant information about the empirical sample. The two companies do have in place a long-standing partnership and were chosen because they showed interest on the topic of collaborative relationship performance measurement and management (Table I).

Five individual interviews were conducted in addition to a roundtable workshop with all the informants. These were held at the Buy-C headquarters and aimed at openly share opinions on the RelReg. Each interview lasts one or two hours, while the roundtable lasted three to four hours. All the interviews have been recorded and transcribed verbatim. Questions included the following: perceived benefits and criticalities of the system; the comprehensiveness of the set of performance dimensions under scrutiny; the feasibility and reliability of the collaborative design, implementation, use and review; opinions on suitable application contexts; and suggestions for improvement and free thoughts on the PMS. The interview questionnaire is reported in Appendix.

4. Framework development and preliminary validation
In this section, the RelReg is described along all the phases of its lifecycle. The first paragraph deals with the preliminary activities managers should take care of before setting a RelReg, selecting the right partners to propose the project and formalize a relationship strategy. The second paragraph reports the RelReg constituent features, highlighting the performance dimensions to tackle and the design process. The third paragraph finally presents the main activities the buyer and the supplier should take care of along with the implementation, use and review phases.

4.1 Antecedents of RelReg adoption
It is a common thought in the operations management literature that competition is no longer between companies, but among SCs, leading to the concept of SC-based competition (Zhang, 2006; Qi et al., 2011). This is a critical concept per se, often treated superficially by referring to misleading formulas like the “supply chain vs supply chain” game. In most industries (e.g., consumer goods, consumer electronics, pharmaceutical, automotive, etc.), competing SCs appear more like interconnected or overlapping networks than mutually exclusive chains of companies enrolled in a tier vs tier competition. Companies are nodes in fuzzy enterprise networks more than tiers in straight SCs: in this context, strategic SCM practices could be exploited in order to create privileged path, thus achieving sustainable competitive advantage. The management of buyer-supplier relationships is therefore
essential for achieving superior performance. Our effort to develop a buyer-supplier collaborative PMS is a concrete attempt to orientate buyer-supplier dyad to increasing collaboration and continuous improvement.

A first aspect to consider is that as SCs become increasingly complex, companies are likely to interact with a lot of external partners. From the RelReg sponsor perspective (either a buyer or a supplier), it is therefore of vital importance to carefully select the right SC partner to engage. A structured approach to portfolio management is therefore a key antecedent to succeed. Strategic relevance of the partners, current relationship capabilities in place, technical feasibility are some of the factors that should be taken into account.

Taking for granted a high commitment from both parties involved, it is then of fundamental

<table>
<thead>
<tr>
<th>Unit of analysis</th>
<th>Supplier (metrics)</th>
<th>Buyer (metrics)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Financial dimension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“To achieve financial value from this relationship, what parameters should be optimized?”</td>
<td>Business relationship</td>
<td>Revenue growth</td>
</tr>
<tr>
<td></td>
<td>Total cost of sales</td>
<td>Total cost of ownership</td>
</tr>
<tr>
<td></td>
<td>Distribution costs</td>
<td></td>
</tr>
<tr>
<td><strong>Operative processes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“To ensure routinely operational excellence, which SC operational activities should be optimized?”</td>
<td>Order cycle</td>
<td>Agreed order fulfillment</td>
</tr>
<tr>
<td></td>
<td>Order</td>
<td>Order fill rate, order lead time</td>
</tr>
<tr>
<td></td>
<td>Delivery process</td>
<td>Punctuality index, flexibility index</td>
</tr>
<tr>
<td></td>
<td>Invoicing</td>
<td>Invoicing accuracy, invoicing timeliness</td>
</tr>
<tr>
<td></td>
<td>Payment</td>
<td>Payment timeliness, documentation accuracy</td>
</tr>
<tr>
<td></td>
<td>New product development</td>
<td>Product development time</td>
</tr>
<tr>
<td></td>
<td>Tracelability and stock control</td>
<td>Inventory level, security stocks level,</td>
</tr>
<tr>
<td><strong>Planning processes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“To achieve superior coordination, which planning process must we excel at?”</td>
<td>Demand planning</td>
<td>Forecast accuracy, forecast variability</td>
</tr>
<tr>
<td></td>
<td>Production planning</td>
<td>Actual versus planned production</td>
</tr>
<tr>
<td></td>
<td>Distribution planning</td>
<td>Changes entity, changes frequency</td>
</tr>
<tr>
<td><strong>Product/service exchanged</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“To add value for the final customer, which quality target should respect the good exchanged?”</td>
<td>Quality-based performance</td>
<td>Quality rate, number of defects</td>
</tr>
<tr>
<td><strong>Relationship intangible capabilities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“To continuously improve our relationships, which capabilities should we develop?”</td>
<td>Social capital</td>
<td>Mutual trust, Goal alignment, Number of meetings, Perceived value of the relationship</td>
</tr>
<tr>
<td></td>
<td>Information capital</td>
<td>Exploitation of collaborative platforms, Digitalization degree, Information quality, Information timeliness</td>
</tr>
</tbody>
</table>

Figure 2. RelReg dashboard: an illustrative example of the O model
importance to define a formal buyer-supplier relationship strategy that the RelReg should operationalize (Kaplan et al., 2010; Hesping and Schiele, 2015). The relationship strategy formalization should be the synthesis of a shaping and reshaping process of the two parties’ own strategies. It is of primary importance that the buyer and the supplier eventually agree on a limited set of strategic objectives, acting as the basement of their relationship.

To conclude developing a RelReg is a game of “focus and choice.” Focus on your urgent SC needs and choose the right SC partner to engage in the project, through a mature portfolio management approach. Then, focus on both parties’ needs within the specific relationship and choose a limited number of agreed goals to pursue.

4.2 The RelReg design features
The essence of the RelReg is to enable collaboration within the buyer-supplier performance measurement and management process. The implicit logic of traditional external SCPMS (Simpson et al., 2002; Kannan and Tan, 2002) can be resumed in the following statement: “autonomously measuring something to evaluate someone.” The RelReg turns this logic into: “joint measuring something to collaborate on mutual performance.” In other words, we shift the logic of the external SCPMS from a tool for evaluation to a tool for collaboration (Giannakis, 2007). To put this statement into practice, we should facilitate the rising of collaboration all along the RelReg lifecycle, starting from its design characteristics. Figure 2 shows the RelReg dashboard.

The metrics reported in the RelReg above are just examples: the set of metrics to select within each category is strictly dependent on the strategic goals of the relationship and the availability of data. The performance category introduced is deriving both from the literature (Simpson et al., 2002; Kannan and Tan, 2002; Luzzini et al., 2014) and from panel of experts’ workshops and interviews joined by the authors A deeper analysis follows, highlighting the core questions animating debate with panel of experts on each dimension:

- Financial dimension: to achieve financial value from the relationship, which parameters should be optimized? Notwithstanding the corporate strategy, profitability is ultimately the key objective of every profit-oriented company. Empirical evidence highlights that financial strategies are simple; companies can make more money by: selling more; and spending less. Any programs put in place (and strategic buyer-supplier partnership are no exception) create more value for the company only if it leads to selling more or spending less. Thus, the company’s financial performance gets improved through two basic approaches – revenue growth and productivity. Considering the buyer-supplier relationship within the RelReg, the buyer would be primarily compelled to lower the total “cost of ownership” of acquiring goods/service from the supplier. The supplier instead would be interested to lower the total cost of sales and to increase the revenues within the specific customer relationship.

<table>
<thead>
<tr>
<th>Company</th>
<th>Role in the industry</th>
<th>Revenues</th>
<th>Employees</th>
<th>Informants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buyer company (Buy-C)</td>
<td>OEM – first tier supplier of car maker (automotive industry)</td>
<td>7.9 mld €</td>
<td>43,000</td>
<td>Purchasing Operations Manager</td>
</tr>
<tr>
<td></td>
<td>Core business: production of electronics system, automotive lighting, suspension systems, etc.</td>
<td></td>
<td></td>
<td>Supplier Quality Manager</td>
</tr>
<tr>
<td>Supplier company (Sup-C)</td>
<td>Second tier supplier (automotive industry)</td>
<td>np</td>
<td>np</td>
<td>Logistics Manager</td>
</tr>
<tr>
<td></td>
<td>Core business: production and distribution of electronics and piezoelectric modules and components for the automotive industry</td>
<td></td>
<td></td>
<td>Sales and Customer Service Manager</td>
</tr>
</tbody>
</table>

Table I. Empirical sample (dyadic case study)
Operational processes: to ensure routinely operational excellence, which SC processes should be optimized? Operational processes sustain the daily flows of materials, information, documentation and money between the buyer and the supplier (Kaplan and Norton, 1996). Depending upon the buyer-supplier strategy, more emphasis could be given to efficiency or effectiveness. Depending upon the buyer-supplier processes in place, the unit of analysis is modeled.

Planning processes: to achieve superior coordination, which planning process must we excel at? Mutual and anticipated visibility of demand, production and distribution plans is at the basis of interface process coordination. Monitoring the accuracy of these plans will stimulate a continuous improvement in the overall planning processes, which could eventually result in operational improvement (Gunasekaran et al., 2001, 2004).

Product/service quality performance: to add value for the final customer, which quality target should respect the product/service exchanged? The overall quality of the product/service supplied is critical to add value along the SC and deliver something appealing for the end consumer (Simpson et al., 2002; Kannan and Tan, 2002).

Relationship intangibles capabilities: to sustain our relationship, which capabilities should we develop? This dimension identifies the intangible assets that are important to stimulate and fuel mutual collaboration. We distinguish between social capital and information capital. The former relies in the degree of integration and mutual trust characterizing the relationship between the two parties. The latter identifies the availability of information system, networks and infrastructure required to support the buyer-supplier strategy. Qualitative metrics based on a Likert-scale questionnaire submitted to both the buyer and the supplier company could be used (Cousins et al., 2008).

Within the five categories identified, we coherently distinguish between metrics addressing the supplier’s performance, metrics addressing the buyer’s performance and transversal metrics jointly addressing both parties. Each metric in the RelReg should be exploded according to the paradigm proposed in the balanced scorecard (Kaplan and Norton, 1996) and re-adapted coherently with the presence of two actors. Figure 3 reports all the information to be defined for each metric.

First, the strategic objective underneath should be stated as well as the precise measures on which the buyer and the supplier have to converge. Then the target should be included, identifying which actor defines it. Finally, a set of initiatives aimed at achieving the target have to be listed: in most cases (especially when dealing with SC operational and planning processes or relationship capabilities), even if a certain performance is evaluating just one actor (either the supplier or the buyer), the relationship partner could also provide its support to improve the performance. The punctuality index as reported in Figure 2 is a good example: the supplier is mainly responsible and should act on his delivery planning process and in the downstream transportation operations to improve the performance. On the other hand, the buyer company could also provide an active support, for example, by improving its forecasts or optimizing the inbound logistics operations. Generally speaking, it is essential that both actors involved jointly manage various steps of the RelReg adoption.
4.3 The RelReg lifecycle

From internal PMS literature, we know that most performance measurement and management projects fail because they are poorly implemented, used and reviewed, rather than poorly designed (Bourne et al., 2003). In order to make the RelReg as a “ready to adopt” tool in the hands of practitioner, it is worth reporting in this research which are the main elements to consider within each phase of the PMS lifecycle and how they should be shaped when applying to a buyer-supplier PMS as the RelReg.

Once the RelReg is designed, it should be implemented. The implementation phase entails all the procedural steps enabling measurements to be made regularly (Bourne et al., 2000, 2002, 2003): data collection and integration, performance measures calculation and reporting management. In the RelReg, these basic steps should be shaped coherently with the presence of two actors. Due to the existence of metrics addressing the buyer and metrics addressing the supplier, the primary data collection naturally involves the two parties. Each part is initially responsible to provide reliable data, either qualitative or quantitative, and to rigorously calculate the performance measures. The reporting phase is particularly important in enabling the collaborative approach on mutual performance, the fundamental logic behind the RelReg. A complete visibility on mutual performance agreed on the RelReg should be allowed. The frequency of the reporting depends upon the industry and the timing of the mutual flows between the buyer and the supplier. The management information systems are critical to the success of PMS implementation (Nudurupati et al., 2011; Ho, 2007), particularly in data collection, analysis, presentation and dissemination (Neely, 1999). In the RelReg case, they should enable an efficient and effective integration of the two parties; MIS technology innovation like web-based or cloud platform for data sharing could be exploited.

The way the PMS is used ultimately defines its purpose and the expected outcome from the adoption. Using a PMS implies activities like feedback management, discussion on performance reported, improvement in plan design, contract and incentives management. The most referred framework describing the PMS use is the diagnostic vs interactive paradigm (Henri, 2006). The diagnostic use reflects a traditional top-down feedback approach. Measures are used to unilaterally track progress toward goals, monitor results, compare outcomes to expectations and drive rewards mechanisms accordingly. The interactive use reflects a bi-directional role of the PMS, which enables discussion on results and fosters continuous improvement while, in the meantime, improving functional integration. The RelReg, as a collaborative buyer-supplier PMS, is more oriented toward an interactive use. However, this should not relax the constant effort on continuous improvement: roles, responsibilities and consequences should be clear from the beginning.

The last step of the PMS lifecycle entails the review of the set of metrics adopted, aimed at keeping the PMS constantly aligned with a changing strategy (Braz et al., 2011). Reviewing the RelReg first implies to detect changes in the relationship strategies and coherently re-formulate the goals of the collaboration. Contextual variables (like technological innovations, changing in customers’ needs, competitors’ actions, new industry regulations, etc.) or company specific (changes in the business strategy, new supplies need) could lead to review the buyer-supplier relationship strategy. Then, coherently with the new goals, the design features of the measurement tool should be updated, by introducing new metrics, or changing the targets and initiatives to existing ones. Figure 4 graphically shows all the steps of the RelReg lifecycle.

4.4 Insights from a dyadic case study

The first insight emerged from the case study is the different reactions among the buyer company and the supplier company in respect to the RelReg. The Sup-C representatives appeared enthusiastic, showing a positive attitude toward the innovative
tool since the beginning. Sup-C does have in place a customer PMS, aimed at recording payment behaviors of customers, yet the measurement is not structured and is never reported to the counterpart. The RelReg is designed to allow a bilateral performance measures sharing, fostering mutual collaboration more than control and evaluation. This element hit the attention and curiosity of the Customer Service Manager: “it happens that customers engage us in performance improvement plans to fix some specific issues. We are always willing to collaborate and improve our performance to better serve the customer, yet it sometimes happens that the problem is within the customer operations, not ours. More often, to maximise a specific relationship performance, a one party’s effort is just not enough.” In his view, since metrics have the power to focus the attention on what really matters, the RelReg would help in freeing the potential of the relationship. Looking at the other side of the coin, the Buy-C representatives appeared more cautious about the tool. Reporting the words of the Purchasing Manager: “we pay the supplier and therefore it seems logical that we set the indicators with relative targets […] however it is
true that with some very important suppliers such a tool could help in maximizing performance.” The Supply Chain Manager further elaborates on this: “I’m interested in the operative and planning processes performances; on both dimensions a mutual visibility is of critical importance […] if I want suppliers to be reactive, flexible and punctual, unilateral service level agreements may not be enough; such a tool could be useful at least in stimulating a more systematic information exchange.”

The interviewees were then asked to comment on the completeness of the RelReg in terms of performance dimensions addressed. The Buy-C representatives agreed the most innovative performance dimensions present to be the financial and relationship soft capabilities dimensions. As the Purchasing Manager admitted: “we already have in place a lot of metrics on the operative processes and on the product/service exchanged […] but, while we agree that the goodness of the relationship is a critical element, we still do not have specific metrics for this, yet everything is left to buyers’ personal experience and perception about the suppliers. Considering the financial dimension, we always look at the suppliers’ financial statements in the supplier selection phase; then the information is not tracked systematically and this has led to unpleasant surprises in the past.” The Supplier Quality Manager adds on that: “One aspect that is missing within the RelReg is the sustainability performance, which is increasingly important for our company and invest our suppliers consequently. Nowadays, as for the financial performance, sustainability is carefully and thoroughly monitored only in the supplier selection phase.” Also, the Sup-C representatives acknowledged the completeness of the RelReg dashboard; in particular, they seemed attracted by a higher control on transactional costs, yet rather skeptical about the feasibility of putting in place a reliable costing system.

As far as the RelReg lifecycle is concerned, everybody agreed: “starting such a thing is the most critical issue!” citing the Buy-C Supply Chain Manager. They highlighted that, though collaboration and mutual trust may be taken for granted within highly collaborative relationships, moving forward toward the RelReg co-design is far from being immediate. The Buy-C purchasing manager highlighted: “set a good PMS is quite a tortuous journey per se […] developing it with a supplier entails a greater effort, though I admit benefits could be higher in the long run.” The specific metric definition (cf. Figure 3) resembles this complexity. Reporting the words of the Sup-C Customer Service Manager added: “I appreciate all the information needed to be collected for the single metric […] often we receive rather obscure metrics from our customers, with no idea on how they have been calculated; this could be frustrating. Considering targets, this is even worse […] sometimes they are simply unreliable. […] The RelReg approach is very mature, though I honestly do not know how many customers within our portfolio would be willing to undergo this process.” The implementation of the RelReg was credited as problematic too, as the data collection process directly involves both actors. Quoting the Buy-C Supply Chain Manager: “the implementation mechanisms should be well oiled […] if a party fails in collecting data, the overall system fails.” The Sup-C Logistics Manager further elaborates on this: “I think it is of primary importance to ground on a dedicated ICT infrastructure, which would make things more efficient and more importantly would guarantee a rigorous implementation of the RelReg.” Both parties agreed on the fact that the RelReg adoption would more likely start from the buyer company and that top management commitment would be decisive to overcome cultural and operational obstacles to RelReg design and implementation. Once put in place, actively using and periodically reviewing the system was considered as less critical. Nonetheless, given the great effort and buyer-supplier integration needed along the RelReg design and implementation, the commitment of both parties in using and reviewing the RelReg in a mature way is somehow taken for granted.

In the final roundtable, Buy-C and Sup-C managers confronted themselves on the main criticalities and benefits of the RelReg adoption. As far as criticalities are concerned,
in addition to the Buy-C cultural reticence in being measured by a supplier, the main issue for both parties was the time and resources needed to adopt the RelReg all along its lifecycle. All the managers strongly remarked how the RelReg could be applied only with a very limited number of SC partners; an upper bound of three parallel RelReg adoption emerged from the round table. Another critical issue raised by the managers concerned the relationship between the RelReg and the other systems in place, i.e. the traditional supplier PMSs (called vendor rating) for Buy-C and the customer evaluation procedures in place at the Sup-C. The Sup-C Customer Service frankly admits: “we do serve several customers and most of them report periodic reports on our performance […] metrics are always different! If we put in place the RelReg, sure we would focus a lot on its performance areas, but we cannot ignore the others! This may further increase the entropy in my department.”

The Buy-C representatives were also worried about the additional entropy brought by the RelReg within an already complex and time-consuming activity as the suppliers’ performance measurement and management. Quoting the Buy-C Supplier Quality Manager: “of course some suppliers are more important than others, but our company quality and service level standards are very demanding and so we can’t neglect the rest of the supply base. Plus, we are always overloaded, so it is important for the RelReg to be as much integrated as possible with our systems and operations.”

On the other hand, both parties emphasize the breakthrough innovation that the tool could bring when applied in a real buyer-supplier relationship. Managers referred to “real” and “systematic” collaboration: real, in light of the bilateral performance measurement; and systematic, because the strategic goals of the relationship are operationalized into a set of shared metrics. Indeed, everybody involved recognized the value added by a structured PMS, under the curtain of the well-known adage “what you measure is what you get.” Both parties eventually recognized that mutual trust in a good relationship is a critical antecedent, otherwise, at the first performance pitfall, things may go wrong and even damage the relationship.

A final issue addressed in the roundtable discussion was the most suitable application context. In addition to a good and long-term-oriented relationship already in place between the buyer and the supplier, the timing for the RelReg adoption emerged as critical. On this behalf, the Purchasing Manager said: “I think the most suitable occasion to start the RelReg adoption would be when we have to involve the designated supplier - one with whom we already have a long standing and trustworthy relationship - into an innovation project which directly involves his supplies. These situations are not rare in our business and could represent the driver for adopting the RelReg.” An agreement among the other manager converges in this point. Thus, the top management commitment was identified as another important ingredient. The Sup-C Customer Service Manager added: “I think that an important innovation in the supply requested could definitely motivate the RelReg, which could sound hard to justify otherwise; moreover, I think the top management commitment to be decisive to overcome obstacles to the adoption.”

5. Discussion

The extant literature does not tackle the issue of collaboratively measuring and managing external SC performance. Nevertheless, this pattern is not diffused within companies: it is generally limited to a short panel of SC actors with partnership relationships and, by the way, not related to a structured performance measurement and management process. Each company within an SC naturally displays a far higher confidence on internally developed SCPMSs than in performance measures coming from external partners. Large companies always measure some kinds of SC performance by themselves and this eventually results in a myriad of metrics flowing within an SC. These metrics often increase the distance among SC partners rather than integrate them (see Purdy et al., 1994; Purdy and Safayeni, 2000).
Focusing on dyadic relationships, each actor generally relies upon its own measures, thus often leading to rigid relationships and to the impossibility to carry on joint performance improvement plans (Hald and Ellegaard, 2011). Performance measures are used to fuel harsh negotiations, increase the bargaining power and develop autonomous local optimization processes (Luzzini et al., 2014). The RelReg challenges this paradigm starting from agreed relationship goals and allowing an active participation of both parties from the design to the review phase. The idea is to adopt the measurement tool to quantify the outcomes of mutual collaboration efforts, thus aiming at continuous improvement and win-win initiatives.

It seems logical for the RelReg to be introduced by the buyer company, once it overcomes internal cultural barriers related to be evaluated by their supplier. Once proposed, we do not expect a negative reaction from the designated supplier, who has nothing to lose from a more structured collaboration. In terms of design features, the managers interviewed mostly recognized the completeness of the tool. On this behalf, the lack of the sustainability dimension was mentioned by the Buy-C Supplier Quality Manager. It may be worth adding it in cases when buyer-supplier relationship strategic goals are strongly related to the sustainability. Nevertheless, we prefer not to further complicate the RelReg basic version dashboard, as sustainability aspects may be included within other dimensions (i.e. operative processes).

As the dyadic case study further remarked, the RelReg adoption is with no doubt a highly time-consuming and resource-demanding process; a partnership should be developed; a joint SCPMS should be designed finding a convergence on the metrics to be adopted; and a reliable management information infrastructure should be put in place to link the two parties and allowing data collection, performance measures’ calculation and reporting. Investments in time and resources are needed, and the benefits could be tangible only in the long run. Another element that clearly emerged from the case study was that while a RelReg tackles a single buyer-supplier strategic relationship, companies interact with a lot of strategic SC partners. Consequently, there may be problems of RelReg integration with other external SCPMSs, both in terms of design features coherency and consistency with systems and implementation procedures. Both parties identify this as a potential source of entropy. Four elements have been highlighted as critical to overcome obstacles and skepticism and allow for a successful RelReg adoption: first, the top management commitment is essential, given the strategic nature of the RelReg in nurturing collaboration within highly strategic buyer-supplier relationship. Second, find the right counterpart: only a very limited subsample of strategic suppliers (cf. Kraljic, 1983) is so critical to justify the RelReg adoption, yet this is not enough: a long standing and trustworthy relationship already in place represents a critical antecedent. Third, the timing of the RelReg introduction appears to be more suitable in supporting an innovation process within the specific buyer-supplier relationship. It should be important to use the RelReg in this phase with a certain degree of flexibility, considering that the goals of a buyer-supplier relationship may undergo some changes. Fourth, maximize the consistency of the RelReg adoption in respect to existent SCPMSs in terms of design features and implementation systems and operations: when possible, it would be intelligent to use for the RelReg the same metrics as the other systems. This will allow for a structured benchmark between highly important suppliers (customers) and the rest of the supply (customer) base.

6. Conclusions
This study reports the RelReg as an innovative buyer-supplier PMS aimed at collaboratively measuring buyer-supplier relationship performance. Both the system design features (the RelReg dashboard) and activities to be performed all along the adoption process (the RelReg lifecycle) are provided. Finally, the RelReg is empirically tested and improved and supported by the dyadic case.
A thorough empirical validation for the proposed framework is an action for future research. While the dyadic case study represents a first step in this direction, the present paper should be considered as a mainly conceptual paper, aimed at proposing an innovative framework for buyer-supplier collaborative performance measurement and management. We therefore encourage both scholars and practitioners to implement the RelReg in a real buyer-supplier relationship, in order to further refine and test the proposed model. Another limitation of the study relies in the link between the buyer-supplier strategy and the RelReg adoption. The measurement tool naturally comes after a mapping of the relationship strategy, aimed at highlighting key goals to operationalize. Future research should tackle the issue on how to build and describe the relationship strategy and consequently shape the RelReg coherently, also challenging the performance dimensions identified, if necessary.

We deem this paper to have several managerial implications. The RelReg is supposed to be a simple and smart tool, ready to be applied by practitioners. Indeed, we have provided normative guidelines (Section 4) that could support companies within the RelReg adoption process. First, we highlighted the importance of portfolio management and of buyer-supplier strategy definition as fundamental antecedents. Once the right counterpart has been chosen and strategic goals have been agreed, the RelReg could be designed by selecting key metrics within the clusters identified. Finally, providing guidelines on the implementation, use and review also, we aim at supporting practitioners all along the steps of the lifecycle.

This study is at the crossroads between the external SCPMS literature and buyer-supplier relationship management literature, and display theoretical contributions to both. Rather than addressing the topic from a holistic perspective like other research before (think about SCOR-based framework), we propose the single dyad as the fundamental unit to tackle (see the case study reported). Then, we build a new measurement tool, which allows both parties to take an active role in the measurement and management process. Our hope is that this study may fuel a new stream of research based on buyer-supplier collaborative performance measurement and management.

References


Further reading


Appendix. Interview questionnaire

RelReg dashboard

- Which are the elements of the RelReg dashboard that most help you to do your job? And why?
- From your perspective, which elements of this dashboard could be considered as the “most important” for you to manage your relationship with the SC partner?
- Do you like the idea behind the RelReg, i.e. providing an innovative buyer-supplier PMS aimed at fostering in a structured way the collaboration on mutual relationship performance?
- Do you think that all relevant performance dimensions are covered (cf. the five blue boxes)? If not, what is lacking?
- What are the most important performance dimensions? Why? Could you mention three examples of that?
- What is the least important performance dimension? Why? Could you give me an example?
- What would you change in the model?
- Do you think that all sub-units within each performance dimension are covered?
- Is the graphical representation effective?
- Do you have any other considerations on the RelReg design features?
- Do you think that this dashboard needs to be shared with other functions/departments in your organization? Why?

Metric definition

- Do you think the amount of information to be tracked within each metric is enough?
- If not, which kind of information is missing?
- Do you think that some piece of information within the metric focus is redundant?
- If yes, which element(s) would you take off?
- Do you have any other consideration related to how to characterize each metric?
RelReg lifecycle

- Do you like the idea of maintaining a high buyer-supplier integration all along the RelReg adoption process (from design to review)?
- Do you think it is feasible?
- Which could be the main barriers?
- Which additional benefits you may recognize? Can you give me three examples please?
- Would you be ready to adopt the RelReg with some SC partners? Why? And when?
- In which industries do you think the RelReg could be more suitable? Why?
- Which are the drivers for the RelReg adoption?

Corresponding author

Vieri Maestrini can be contacted at: vieri.maestrini@polimi.it
The impact of supplier performance measurement systems on supplier performance

A dyadic lifecycle perspective

Vieri Maestrini
Department of Management, Economics and Industrial Engineering,
School of Management of Politecnico di Milano, Politecnico di Milano, Milan, Italy

Davide Luzzini
Department of Marketing, Operations and Supply,
Escuela de Alta Dirección y Administración, Barcelona, Spain and
Department of Information System and Supply Chain Management,
Audencia Business School, Nantes, France, and

Federico Caniato, Paolo Maccarrone and Stefano Ronchi
Department of Management, Economics and Industrial Engineering,
School of Management of Politecnico di Milano, Politecnico di Milano, Milan, Italy

Abstract

Purpose – The purpose of this paper is to empirically investigate the impact of a mature supplier performance measurement system (SPMS) adoption all along its lifecycle phases (i.e. design, implementation, use and review) on the suppliers’ performance.

Design/methodology/approach – The research hypotheses have been tested on a final sample of 147 pairs of buyer-supplier responses, collected by means of a dyadic survey involving manufacturing firms and one key supplier of their choice. The research framework has been tested through a structural model using PLS regression.

Findings – Considering the joint effect of all the four SPMS phases on supplier performance, the findings show that the system use and review play a prominent effect: the former have a positive impact on supplier quality, delivery and sustainability performance; the latter positively affects supplier delivery, innovation and sustainability. A mature design displays a positive effect on supplier sustainability performance, while a mature implementation results to negatively affect supplier innovation performance. Finally, cost performance is not impacted by any of the four phases.

Originality/value – This study contributes to the open debate regarding the relationship between SPMSs and actual supplier performance improvement. In particular, the lifecycle perspective is introduced to clearly distinguish among each phase of adoption and assess their relative impact on supplier performance. Besides, the dyadic nature of the study allows to investigate different subcomponents of supplier performance jointly considering the buyer company and supplier company perspective, thus achieving a more insightful and robust information.

Keywords Performance measurement, Survey, Buyer-supplier relationships

Paper type Research paper

1. Introduction

Suppliers today are increasingly responsible for buyer companies’ value creation. For this reason, the careful assessment and monitoring of performance upstream in the supply chain is critical. Hence, supplier performance measurement systems (SPMSs) are critical tools that managers exploit to control and orchestrate suppliers.

The academic literature on SPMSs is quite extensive, though it is mostly focused on the SPMS design. Indeed, most contributions answer the “what to measure” question, focusing on metrics selection and composition into measurement frameworks (Simpson et al., 2002;
Kannan and Tan, 2002). Moreover, the impact of SPMSs on supplier performance has been poorly investigated. While a few contributions have addressed the performance outcomes of SPMS design features (Carr and Pearson, 1999; Mahama, 2006; Cousins et al., 2008) and implementation-related aspects (Prahinski and Benton, 2004; Prahinski and Fan, 2007), the use and review of SPMSs have been neglected so far.

In this paper, a lifecycle perspective is applied to SPMSs (Bititci et al., 2006; Gutierrez et al., 2015), introducing the concept of maturity of the following four phases of the SPMS lifecycle: design, implementation, use and review. In light of resource orchestration theory (ROT) (Sirmon et al., 2011), we expect the SPMS adoption throughout its lifecycle to have a positive impact on suppliers’ performance (i.e. quality, delivery, innovation, sustainability and cost). ROT has been advanced in conjunction with the resource-based view (RBV) of the firm (Hitt et al., 2011) and states that firms can achieve a sustainable competitive advantage by orchestrating critical resources at their disposal through stages of structuring, bundling and leveraging. ROT appears to fit well the context of buyer-supplier relationship management, by considering suppliers as key resources that the buyer company needs to properly orchestrate. The SPMS is a key tool by which this orchestration is guaranteed, ultimately leading to superior supplier performance. Hypotheses are coherently developed assuming a positive relationship between a mature SPMS adoption and suppliers’ performance (i.e. quality, delivery, innovation, sustainability and cost). To test such hypotheses, this research applies a dyadic perspective by triangulating responses regarding supplier performance from both the buyer company and a key supplier.

The remainder of the paper is organized as follows. In the next sections, ROT is presented and the literature on PMS lifecycle is reviewed to develop measures for the SPMS lifecycle maturity. Then we report the research framework and the hypotheses to be tested, and explain the method adopted. In the remaining sections, findings of both the measurement and the structural model are reported and then discussed. Conclusions section end the paper.

2. SPMS and resource orchestration

A PMS is defined as a set of metrics used to quantify the efficiency and effectiveness of actions (Neely et al., 1995). Similarly, an SPMS is defined as a set of metrics used to quantify the efficiency and effectiveness of suppliers’ actions (Hald and Ellegaard, 2011; Maestrini et al., 2017). As the main goal of a PMS is to support the implementation of strategy at various levels (Kaplan and Norton, 1996), the goal of an SPMS is to support the purchasing strategy and align buyer-supplier relationships toward consistent goals (Kannan and Tan, 2002).

The RBV has long claimed that a sustainable competitive advantage is derived from owning bundles of valuable, rare, inimitable and non-substitutable resources (Hitt et al., 2011; Wowak et al., 2013). Based on this nuance, Hansen et al. (2004, p. 1280) argue that “what a firm does with its resources is at least as important as which resources it possesses.” Further elaborating on this concept, Sirmon et al. (2011) advance the argument that while owning the right resources is essential, competitive advantage comes from the ability of a firm to “orchestrate” its resources. Indeed, only with a proper resources orchestration, a company manages to execute its strategy. The idea of resource orchestration grounds on the seminal works of Sirmon et al. (2007) and Helfat et al. (2007).

ROT complements RBV in that it explains how resources are transformed into capabilities. According to Sirmon et al. (2007, p. 273), resource orchestration entails “the comprehensive process of structuring the firm’s resource portfolio, bundling the resources to build capabilities, and leveraging those capabilities with the purpose of creating and maintaining value for customers.” An in-depth discussion of these three stages is
outside the scope of this study. What matters to us is extending the application of ROT to the context of buyer-supplier relationships and specifically to the role of the SPMS for resource orchestration.

As a matter of fact, we know that resources can be both internal and external to the firm. In particular, purchasing and supply management can yield competitive advantage *per se* (as an internal capability that is unique to the firm) and allowing the firm to access suppliers’ resources and capabilities (Barney, 2012). ROT seems particularly suitable to interpret the buyer-supplier relationships: structuring the supply base refers to the buyer company’s effort of maintaining a current and aligned portfolio of suppliers. Bundling supplier resources requires allocating specific relational investments with each supplier to access and develop their resources and capabilities. Finally, the buyer company can leverage these resources and capabilities by selecting and activating the right set of suppliers for the achievement of its goals.

Throughout this process, the role of management is empowered. Managers are responsible for setting the vision, deciding upon resource allocation and acting to stimulate resource mobilization (Crook *et al.*, 2008; Hitt *et al.*, 2011). This includes organizing the resource portfolio, developing capabilities out of resources and acting on them to create value (Sirmon and Hitt, 2003). To this end, SPMSs are key tools managers can rely upon to orchestrate supplier resources. Indeed, the SPMS supports the buyer company in each step of the process. By collecting information and measuring supplier performance, the buyer company is able to structure the supply base according to its needs, investing on those that are in line with the company strategy and divesting from those that are not. The SPMS also helps bundling supplier resources, as it allows identifying suppliers that have the potential to generate a competitive advantage and to develop such potential through appropriate supplier development programs. Finally, through the SPMS, the buyer company can mobilize the desired sets of suppliers in accordance to the company strategy.

All in all, the overarching premise of this study is that the SPMS maturity all along its lifecycle (design, implementation, use and review) can grow the buyer company orchestration capabilities and lead to higher supplier performance. In the next sections, we will provide an overview of the SPMS lifecycle stages and their expected link with supplier performance.

3. SPMS lifecycle
The subsequent four paragraphs address the phases of PMS design, implementation, use and review, identifying the features characterizing the maturity for each phase. Such phases are mostly grounded in the PMS literature and serve as the premise to introduce the SPMS lifecycle framework which is at the basis of this study.

3.1 PMS design
Designing the PMS consists of the identification of key objectives to achieve, derived from the company strategy, with their subsequent operationalization into a set of metrics (Neely *et al.*, 1995). A proper metric requires the following three distinctive elements: a performance measure that quantifies what is happening; a performance standard, or target, that discriminates between good and bad performance; and consequences related to being on, below or above target (Melnyk *et al.*, 2014).

The alignment of the PMS with the corresponding strategy is a critical element that should be guaranteed in the PMS design phase by including all performance dimensions at stake (Chenhall, 2005; Franco-Santos *et al.*, 2012). Another critical element for a successful PMS design is the involvement of all the relevant stakeholders in the process (Choi *et al.*, 2012; Papalexandris *et al.*, 2004; Sandstrom and Toivanen, 2002).
Considering the SPMS, the design phase entails the operationalization of the goals related to supplier relationship management into a set of metrics, addressing critical suppliers’ performance (Luzzini et al., 2014). Most diffused performance dimensions monitored are quality, delivery and cost performance (Kannan and Tan, 2002; Gunasekaran et al., 2004). More advanced systems tend to also include metrics measuring supplier capabilities in terms of innovation effort and sustainability approach (Kannan and Tan, 2002; Simpson et al., 2002). Within the SPMS lifecycle, the design phase is the most debated in scientific literature: contributions span from metrics selection procedures (Igarashi et al., 2013; Masella and Rangone, 2000; Huang and Keskar, 2007) and innovative measurement framework presentation (Humphreys et al., 2007; Carter et al., 2010; Muralidharan et al., 2002) to the outcomes of SPMS design (Mahama, 2006; Cousins et al., 2008).

In accordance with the internal PMS literature, the following three main elements can be used when evaluating the maturity of the SPMS design: the metric set completeness, which addresses the presence of all the critical supplier performance dimensions (Kannan and Tan, 2002; Simpson et al., 2002); the involvement of all the relevant stakeholders in the design process (e.g. organizational functions other than the one owning the SPMS and/or the suppliers themselves – Luzzini et al., 2014); and the presence of a robust alignment with the purchasing/supply chain (SC) and eventually the business strategy (Gutierrez et al., 2015).

3.2 PMS implementation
The implementation phase of the PMS consists of establishing procedures and systems to collect, analyze and disseminate data, enabling regular measurements (Bourne et al., 2000; Lohman et al., 2004; Garengo et al., 2007). It is characterized by the activities of data collection and collation, metrics computation and reporting processes (Bourne, 2005; Bourne et al., 2003). To efficiently and effectively manage the PMS implementation, an appropriate information and communication technology (ICT) infrastructure is needed, allowing automation and data reliability (Garengo et al., 2007; Nudurupati et al., 2011; Bititci et al., 2006). A poor supporting ICT may negatively affect the PMS reliability and generates mistrust in metrics reported, thus preventing a successful implementation of the tool (Cavalluzzo and Ittner, 2004).

Considering SPMSs, the activities involved are the same (data collection, metrics calculation and system reporting) but are further complicated by the need to collect data from external sources and to manage inter-company reporting. The maturity of the SPMS implementation is defined by the reliability of data collection and metrics’ computation; the presence of a mature ICT allowing for automation in data analysis and metrics calculation; and structured, formal and frequent reporting (Bourne et al., 2000, 2003; Luzzini et al., 2014; Leeuw and van den Berg, 2011).

3.3 PMS use
The PMS use concerns the way the measuring part manages the relationship with the measured part through the PMS. It entails activities such as communication and feedback management, discussion of reported performance, launch of improvement plans, contract and incentives/disincentives management when present (Gutierrez et al., 2015; Hall, 2008; Grafton et al., 2010).

In the present research, as in most operations and accounting literature, the reference model to frame the PMS use is of Henri (2006), who identifies, grounding on the previous work of Simons (1995), the following two main paradigms of PMS use: diagnostic and interactive. The former entails a traditional mechanistic control, carried out in a top-down fashion by the measuring party (i.e. the buyer company) toward the measured party (i.e. the supplier); it is based on formal reporting and constant control of target achievement. The latter enables the use of PMS to stimulate mutual dialogue and open discussion on
reported performance, with the goal of continuous improvement by enhancing collaboration among the parties involved. According to Henri (2006), these two approaches could co-exist, leading to dynamic tension, which arises from the combined use of the PMS in a diagnostic and interactive fashion. In this way, the two methods of use can be seen as complementary forces, jointly shaping the performance management process. Empirical evidence from past studies on PMS use shows how this dynamic tension actually displays the most positive impact on performance (Widener, 2007; Mundy, 2010; Koufteros et al., 2014).

In the literature on supplier performance measurement, a focus on SPMS use is still lacking; yet, it is critical to depict the buyer company approach in measuring and managing the supplier performance. In this paper, SPMS use is framed by transposing the interactive vs diagnostic framework to SPMS and buyer-supplier relationship management. Doing so, the diagnostic and interactive use of SPMS can be identified. The former allows for monitoring the supplier coherently with performance measures and targets set by the buyer company; the SPMS in this case is mainly exploited to align supplier behavior to the buyer purchasing strategy. The latter entails an active supplier involvement in the measurement process, emphasizing the bi-directional nature of the relationship; the SPMS facilitates dialogue and open debate on mutual performance, aiming at continuous improvement through win-win performance improvement plans. On the one hand, a diagnostic SPMS use is cost and time efficient, but it could result in stiffening the relationship between the buyer and the supplier. On the other hand, an interactive SPMS stimulates a positive climate fostering collaboration, but is more difficult to apply and more demanding in terms of time and resources. Following Henri (2006), the coexistence of these two approaches leads to a dynamic tension between the diagnostic and interactive use of the SPMS. Coherent with internal PMS literature, buyer companies are expected to gain the most from this dynamic tension since suppliers are supported and stimulated in a positive manner while constantly controlled and directed towards target achievement. Within this paper, combined high levels of interactive and diagnostic features (dynamic tension) are linked with a mature use of the SPMS.

3.4 PMS review
The PMS review is the last phase of the PMS lifecycle and involves the PMS revision by updating targets and introducing new performance measures, in order to assure alignment with the strategy over time (Kennerley and Neely, 2003; Bourne et al., 2000; Braz et al., 2011). Although it is often neglected by companies and poorly discussed in scientific literature, timely reviewing of the PMS is of crucial importance. If changes in the strategy are not reflected in the PMS, the alignment is lost and the PMS orchestrating role is not coherent with the strategic goals.

Considering SPMSs, the review is even more critical because purchasing strategy and goals are much more volatile and dynamic than business strategy (Hesping and Schiele, 2015). Thus, constant attention to SPMS review is critical. Therefore, the maturity of the SPMS review phase is identified by the degree of the introduction of new metrics and the frequency of target updates (Braz et al., 2011; Bourne et al., 2000).

4. Research framework and hypotheses development
Sirmon et al. (2007, 2011) claimed that resource orchestration requires the following: structuring the resource portfolio (i.e. acquiring, accumulating and divesting resources), building resources (i.e. establishing current capabilities and developing new ones) and leveraging resources (mobilizing and coordinating capabilities). PMSs facilitate this type of orchestration (Mehnyk et al., 2004).

Since suppliers represent critical resources for the buyer company, purchasing/SC managers need to properly orchestrate them. Coherent with ROT, the main argument
advanced in this paper is that a mature SPMS adoption, along its whole lifecycle, guarantees a successful suppliers’ orchestration. Indeed, an SPMS enables supplier portfolio management, the development of supplier capabilities and, ultimately, the coordination of suppliers, aligning their actions with the buyer company’s needs. Thus, our theoretical framework reports a direct positive relationship between a mature SPMS lifecycle and suppliers’ performance in terms of quality, delivery, innovation, sustainability and cost (see Figure 1).

The following paragraphs discuss the relationship between each SPMS lifecycle phase and supplier performance, reporting the hypotheses of the model.

4.1 The relationship between SPMS design maturity and supplier performance

Within SPMS literature, several survey-based works highlight a positive relationship between the adoption of different performance metrics and performance. Mahama (2006) found a positive relationship between the adoption of both financial and non-financial metrics and supplier performance in terms of quality, savings, on-time delivery and improved decision making. Cousins et al. (2008) found a positive relationship between the inclusion of both “operational” metrics (delivery to schedule, lead time and conformance to specifications) and “communication” metrics (communication effectiveness, information quality, and regularity of feedback) and buyer business performance, mediated by socialization mechanism. Other case-based studies examined the design process, highlighting the importance of aligning the SPMS with pre-defined relationship goals and to involve all relevant stakeholders, starting from other organizational functions within the buyer company (Luzzini et al. 2014; Hald and Ellegaard, 2011).

This study advances previous ones by addressing a more complete construct of design maturity, including the presence of several performance dimensions within the measurement framework and the maturity of framework design procedures. First, a mature design should guarantee the robustness and completeness of the measurement framework. Second, in line with ROT, a multi-dimensional SPMS should stimulate the supplier to focus on all the critical performance, leading to a wider positive effect.

Figure 1. Research framework
Thus, the first hypothesis concerns the existence of a positive relationship between the SPMS design maturity and supplier performance, with sub-hypotheses for each performance dimension:

$H1$. The SPMS design maturity positively affects the supplier’s performance.

$H1a$. The SPMS design maturity positively affects the supplier’s quality performance.

$H1b$. The SPMS design maturity positively affects the supplier’s delivery performance.

$H1c$. The SPMS design maturity positively affects the supplier’s innovation performance.

$H1d$. The SPMS design maturity positively affects the supplier’s sustainability performance.

$H1e$. The SPMS design maturity positively affects the supplier’s cost performance.

4.2 The relationship between SPMS implementation maturity and supplier performance

Starting from the early 2000s, some authors within internal PMS literature began to argue that many PMS adoption projects actually fail, not because the system is poorly designed, but rather because it is poorly implemented (Bourne et al., 2000; Bourne et al., 2002). Implementation failures could be due to three main reasons. The first reason is weak data processing along primary data collection, data analysis and metrics calculation; a reliable ICT infrastructure could limit or avoid this problem, along with a clear definition of roles and responsibilities. The second reason is inappropriate content and frequency of reporting, which should be set coherently with the business context, the relationship between the measuring and the measured parties and the supporting ICT. Third, the lack of top management commitment, which may eventually lead to failures in launching the system (Bourne, 2005).

A few contributions in the literature addressed SPMS implementation aspects. Prahinski and Benton (2004) found that, in the presence of a supplier commitment, a structured and formal reporting can improve supplier performance. Prahinski and Fan (2007) focused on reporting frequency, finding that while a high frequency of operational metrics (quality, delivery) reporting positively affects communication quality perception, no significant relationship was found dealing with financial metrics and more strategic-oriented performance.

Reliable data processing from data collection to metrics calculation contribute to a rigorous performance measurement process, which is a critical precondition to establish trust in the system (Hald and Ellegaard, 2011; Purdy and Safayeni, 2000). According to ROT, frequent and structured reporting should empower the orchestration process, stimulating suppliers’ attention toward the progression of performance tracking. Both factors concur with the SPMS implementation maturity. Thus, the second hypothesis can be formulated as follows:

$H2$. The SPMS implementation maturity positively affects the supplier’s performance.

$H2a$. The SPMS implementation maturity positively affects the supplier’s quality performance.

$H2b$. The SPMS implementation maturity positively affects the supplier’s delivery performance.

$H2c$. The SPMS implementation maturity positively affects the supplier’s innovation performance.

$H2d$. The SPMS implementation maturity positively affects the supplier’s sustainability performance.

$H2e$. The SPMS implementation maturity positively affects the supplier’s cost performance.
4.3 The relationship between SPMS use maturity and supplier performance

The diagnostic vs interactive framework (Henri, 2006) has become the reference paradigm for SPMS use. Although related constructs have been operationalized in slightly different ways over the years, most empirical evidence suggests that the best impact on performance is achieved when the diagnostic and the interactive component co-exist, leading to dynamic tension between the two approaches (Henri, 2006; Widener, 2007; Mundy, 2010; Koufteros et al., 2014).

Though some studies approach the way buyer companies use SPMS (Hald and Ellegaard, 2011; Luzzini et al., 2014), the use maturity is a novel construct introduced in this research, corresponding to the diagnostic-interactive dynamic tension. Coherently with ROT, SPMS use directly shapes the supplier orchestration process. More operational and short-term performance dimensions (like delivery, quality and cost) should be positively impacted by a diagnostic use of the system, since they are frequently tracked and easier to measure and quantify. Other more strategic aspects, related to supplier capabilities in terms of innovation and sustainability, may benefit from a more interactive SPMS use, allowing the use of the tool for supplier development. Thus, the third hypothesis is formulated as follows:

H3. SPMS use maturity positively affects the supplier’s performance.

H3a. SPMS use maturity positively affects the supplier’s quality performance.

H3b. SPMS use maturity positively affects the supplier’s delivery performance.

H3c. SPMS use maturity positively affects the supplier’s innovation performance.

H3d. SPMS use maturity positively affects the supplier’s sustainability performance.

H3e. SPMS use maturity positively affects the supplier’s cost performance.

4.4 The relationship between SPMS review maturity and supplier performance

The PMS review has been neglected so far, even dealing with internal PMS. Nevertheless, it plays a critical role within performance management: first it prevents “strategy ossification” (Micheli and Manzoni, 2010), i.e. a rigid and obsolete PMS misaligned with strategic goals. Second, it allows for continuous improvement, since both performance measures and targets are challenged over time (Braz et al., 2011; Gutierrez et al., 2015).

SPMS literature is still lacking contributions on the review phase, though buyer-supplier relationship goals are likely to change over time for a variety of reasons, including changes in the purchasing strategy of a buyer company, relationship disruptions, major market changes or technological innovation. Thus, coherently adjusting the SPMS in a timely manner is important. In accordance with ROT, the orchestration needs of the buyer company may change over time within a specific supplier relationship. Reviewing the SPMS is instrumental in keeping it effective and aligned with relationship goals. Thus, the fourth hypothesis is reported as follows:

H4. SPMS review maturity positively affects the supplier’s performance.

H4a. SPMS review maturity positively affects the supplier’s quality performance.

H4b. SPMS review maturity positively affects the supplier’s delivery performance.

H4c. SPMS review maturity positively affects the supplier’s innovation performance.

H4d. SPMS review maturity positively affects the supplier’s sustainability performance.

H4e. SPMS review maturity positively affects the supplier’s cost performance.
5. Methodology

Given that we rely on research hypotheses that are formulated based on extant literature and that require to be tested on a large sample, we chose to distribute a survey that would allow measuring the SPMS maturity and the supplier performance as a series of latent variables. In particular, data were collected through a dyadic survey process addressing buyer-supplier dyads, in order to solve perception biases (e.g., Aminoff and Tanskanen, 2013; Oosterhuis et al., 2013; Ambrose et al., 2010; Barnes et al., 2006; Kim et al., 1999). The dyadic approach represents an important methodological contribution in respect to previous survey-based single-respondent studies within the SPMS literature (Carr and Pearson, 1999; Mahama, 2006; Heide et al., 2007; Prahinski and Fan, 2007; Cousins et al., 2008). The first paragraph explains the procedures followed to design the survey, select the sample and collect the data. The second paragraph reports the relevant information regarding the constructs’ measures.

5.1 Survey development, sampling, data collection

Two specular versions of the questionnaire (one for buyers and one for suppliers) were prepared, both in English and in Italian. The English versions of the two questionnaires were first developed from literature-driven constructs. Then, they were subjected to a pilot test in order to assure clarity, conciseness and effectiveness in addressing the concepts behind the questions. Both questionnaires were submitted to four English scholars, experts in the field of performance measurement and management, and questions were adjusted based on their advice. The questionnaires were then translated into Italian through the “translation, review, adjudication, pre-testing, and documentation” procedure in order to assure the content’s validity (Harkness et al., 2004). The Italian versions were then submitted to three Italian scholars and to four practitioners (two couples of respondents from two buyer-supplier dyads) for pre-testing. At each step, the wording was adjusted based on the feedback received. Finally, a personalized link to access and fill in the online survey was created to be sent to each company involved.

Buyer companies were addressed first, starting from a population of Italian manufacturing companies with at least 100 employees (smaller companies rarely have a structured SPMS in place). The choice to focus only on the manufacturing sector (ISIC codes from 10 to 33), was primarily aimed at increasing the homogeneity of the empirical sample. Moreover, manufacturing companies are expected to rely on more mature SPMSs, given the historical relevance of suppliers in this sector. No specific requirement was set in advance in respect to suppliers: they were selected by the buyer company’s respondent among the most important active suppliers in terms of spending.

Previous criteria were followed to perform the sampling starting from the AIDA-Bureau Van Dick database (aida.bvdinfo.com) of Italian companies. Buyer companies satisfying the previous requirements were first contacted by phone in order to understand their willingness to take part in the research and were then given instructions on how to fill in the questionnaire. An e-mail with a personalized link to the online survey was sent to those agreeing to participate. The telephone contact and the e-mail text followed a default script developed at the beginning. When filling in the questionnaire, the buyer company respondent had to refer to the selected supplier and provide its contact details. Then, the indicated supplier was contacted following the same procedure used for buyer companies: they were first contacted by phone and then sent an e-mail with the link to the online survey (supplier version). After a defined time of non-response, reminder e-mails were sent.

Buyer companies’ employees tasked with completing the questionnaire were mostly executives in the purchasing or SC functions, knowledgeable about the SPMS in place and about the relationship with the selected suppliers. Supplier respondents ranged from sales to customer service positions. From the total number of buyer companies contacted,
458 agreed to participate in the research, and 238 of them began filling in the questionnaire. The threshold for considering a questionnaire to be acceptable was established as 75 percent of questions were answered. According to this criterion, 204 buyer questionnaires were ultimately usable, leading to a response rate of 44.5 percent. Regarding suppliers, 156 questionnaires were considered complete (according to the threshold), achieving a response rate of 65.6 percent with respect to all the contacts provided by the buyer companies. Finally, matching together buyer-supplier questionnaires, 147 dyads were ultimately usable, meaning that both questionnaires satisfied the completion threshold requisite. Some descriptive information on the usable sample are provided in Table I (buyer companies) and Table II (supplier companies).

After the collection process, the data were cleaned and checked for response bias (Armstrong and Overton, 1977). Non-response bias was tested by ruling out the differences

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Frequency</th>
<th>%</th>
<th>Descriptive</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenues (million €)</td>
<td></td>
<td></td>
<td>Employees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>0</td>
<td>0</td>
<td>Small (1-49)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5-50</td>
<td>43</td>
<td>29.25</td>
<td>Small-medium (50-99)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50-100</td>
<td>30</td>
<td>20.41</td>
<td>Medium (100-249)</td>
<td>62</td>
<td>42.18</td>
</tr>
<tr>
<td>100-1,000</td>
<td>67</td>
<td>45.58</td>
<td>Medium large (250-499)</td>
<td>32</td>
<td>21.77</td>
</tr>
<tr>
<td>≥1,000</td>
<td>7</td>
<td>4.76</td>
<td>Large (500-999)</td>
<td>31</td>
<td>21.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Very large (≥1000)</td>
<td>22</td>
<td>14.97</td>
</tr>
<tr>
<td>Industry sector</td>
<td></td>
<td></td>
<td>Respondent organizational function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machinery and equipment</td>
<td>62</td>
<td>42.18</td>
<td>Purchasing</td>
<td>114</td>
<td>77.55</td>
</tr>
<tr>
<td>Metallurgy and steel goods</td>
<td>17</td>
<td>16.33</td>
<td>Supply chain and logistics</td>
<td>20</td>
<td>13.61</td>
</tr>
<tr>
<td>Chemical and pharmaceutical</td>
<td>15</td>
<td>11.56</td>
<td>Operations</td>
<td>9</td>
<td>6.12</td>
</tr>
<tr>
<td>Textile</td>
<td>11</td>
<td>10.20</td>
<td>Other</td>
<td>4</td>
<td>2.72</td>
</tr>
<tr>
<td>Vehicles</td>
<td>10</td>
<td>7.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food and beverages</td>
<td>8</td>
<td>6.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other manufacturing</td>
<td>24</td>
<td>5.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>147</td>
<td>100</td>
<td></td>
<td>147</td>
<td>100</td>
</tr>
</tbody>
</table>

Table I. Buyer company sample descriptives

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Frequency</th>
<th>%</th>
<th>Descriptive</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenues (million €)</td>
<td></td>
<td></td>
<td>Employees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>28</td>
<td>19.31</td>
<td>Small (1-49)</td>
<td>65</td>
<td>45.14</td>
</tr>
<tr>
<td>5-50</td>
<td>67</td>
<td>46.21</td>
<td>Small-medium (50-99)</td>
<td>22</td>
<td>15.26</td>
</tr>
<tr>
<td>50-100</td>
<td>17</td>
<td>11.72</td>
<td>Medium (100-249)</td>
<td>24</td>
<td>16.67</td>
</tr>
<tr>
<td>100-1,000</td>
<td>25</td>
<td>17.24</td>
<td>Medium large (250-499)</td>
<td>15</td>
<td>10.42</td>
</tr>
<tr>
<td>&gt;1,000</td>
<td>8</td>
<td>5.52</td>
<td>Large (500-999)</td>
<td>5</td>
<td>3.47</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>–</td>
<td>Very large (&gt;1,000)</td>
<td>13</td>
<td>9.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Missing</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Industry sector</td>
<td></td>
<td></td>
<td>Respondent organizational function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metallurgy and steel goods</td>
<td>40</td>
<td>27.2</td>
<td>Sales and marketing</td>
<td>87</td>
<td>59</td>
</tr>
<tr>
<td>Machinery and equipment</td>
<td>27</td>
<td>18.4</td>
<td>Operations</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Distributors</td>
<td>19</td>
<td>12.9</td>
<td>Customer service</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Chemical, iron and steel</td>
<td>14</td>
<td>9.5</td>
<td>Quality</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Wood and paper</td>
<td>12</td>
<td>8.2</td>
<td>Accounting</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Other services</td>
<td>18</td>
<td>12.2</td>
<td>Other</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>Other manufacturing</td>
<td>17</td>
<td>11.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>147</td>
<td>100</td>
<td></td>
<td>147</td>
<td>100</td>
</tr>
</tbody>
</table>

Table II. Supplier company sample descriptives
in terms of size and industry distributions between respondents and non-respondents. Similarly, early response bias was also tested. Both tests showed no significant differences between groups.

5.2 Measures
Constructs were operationalized on the basis of both existing measures within the SPMS literature and measures adapted from other literature streams, in particular internal PMS contributions. Respondents were asked to answer each question on a Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”).

Starting from the independent variable, the SPMS lifecycle maturity has been split into four constructs, addressing maturity at each phase. The SPMS design maturity includes the following three main features: SPMS framework completeness (Cousins et al., 2008; Mahama, 2006; Kannan and Tan, 2002), involvement of other organizational functions (Luzzini et al., 2014; Leeuwen and van den Berg, 2011) and alignment with SC and purchasing strategy (Taylor and Taylor, 2013; Chenhall, 2005). The SPMS implementation maturity is a combination of reliable and rigorous data collection and metrics calculation (Bourne et al., 2000; Grudinschi et al., 2014), automation degree and solid ICT infrastructure (Bourne et al., 2000, 2003; Nudurupati et al., 2011; Bititci et al., 2006), frequent and formal reporting (Bourne, 2005; Leeuwen and van den Berg, 2011; Taylor and Taylor, 2013), and procedures for system implementation (Leeuwen and van den Berg, 2011). The SPMS use maturity is identified with dynamic tension (Henri, 2006; Koufteros et al., 2014) derived from a combined use of the PMS in a diagnostic and interactive fashion. The review maturity is measured looking at the review effort, which is aimed at constantly improving the clarity, usefulness and quality of the SPMS and at keeping the SPMS aligned with the strategy and the changing external environment (Braz et al., 2011; Gutierrez et al., 2015). All these constructs were measured, addressing the buyer side only, considering the 147 buyer responses belonging to a complete dyad; in other words, the buyer who filled out the questionnaire had a linked supplier who responded as well. This choice was made because some lifecycle phases are rather objective and hardly allow for biases (design and review), while others entail activities that are mainly under the scope of the buyer company and not visible to the supplier (implementation).

Addressing the dependent variable, supplier performance was measured by asking the performance impact of the relationship with the specific buyer company and the SPMS in place. Performance dimensions addressed included supplier quality, delivery, innovation, sustainability and cost based on existing scales (Cheung et al., 2010; Gonzalez-Benito, 2007). These constructs were measured considering the mean between the buyer and the supplier responses for each item, exploiting the dyadic nature of the questionnaire. Therefore, the 147 dyadic measures were available to measure the supplier’s performance.

6. Findings
The research model was tested using the partial least square approach (Oh et al., 2012) in Smart PLS. The results of the measurement model, based on a confirmatory factor analysis, are reported in the first paragraph. The results of the structural model are reported in the second paragraph.

6.1 Measurement model
The measurement model counts 42 items asked to buyer companies and 12 asked to supplier companies (responses on the SPMS lifecycle were collected only from the buyer side, while responses on supplier performance were collected from both parties). Overall, 13 multi-item constructs were generated. The SPMS design maturity is a second-order formative
construct that includes completeness of the measurement framework (presence of operational metrics and presence of strategic metrics) and maturity of the measurement process, which are reflective constructs. The SPMS implementation maturity is a second-order formative construct that includes data collection reliability and frequency of reporting, which are reflective constructs.

The SPMS use maturity reflects a dynamic tension between a diagnostic and an interactive SPMS use and is operationalized by means of a second-order construct given by the product of SPMS diagnostic use and SPMS interactive use. The others (i.e. review maturity and various supplier performance) are first-order constructs.

Table III shows the result of the confirmatory factor analysis performed with PLS. The constructs’ validity was verified, first demonstrating evidence for convergent validity. In accordance with Fornell and Larcker (1981) and Nunnally and Bernstein (1994), we checked first-order constructs composite reliability (CR) and the average variance extracted (AVE) to respect the relative thresholds, 0.7 and 0.5, respectively. Discriminant validity was then tested in two ways. The correlation matrix proved that in most cases, the AVE was greater than the square correlation between each pair of latent constructs (Fornell and Larcker, 1981), as shown in Table IV. The heterotrait-monotrait ratio (HTMT) (Henseler et al., 2014) showed good discriminant validity properties with most values lower than the threshold of 0.85 (Table V).

6.2 Structural model
To test the hypotheses of the framework, we ran a path analysis following the suggestions of Peng and Lai (2012) for the use of PLS. The results are shown in Table VI, including standardized path coefficients with two-tailed t-tests for the hypotheses. H1, H3 and H4 are partially confirmed with different levels of significance for the different supplier’s performance, while H2 is not confirmed. Design maturity has a positive impact on sustainability performance ($\beta = 0.366$). Implementation maturity negatively affects innovation performance ($\beta = -0.258$), while not displaying any significant effect on any other performance. Use maturity displays a positive effect on quality ($\beta = 0.271$), delivery ($\beta = 0.225$) and innovation ($\beta = 0.216$) performance. Finally, review maturity positively impacts delivery ($\beta = 0.249$), innovation ($\beta = 0.260$) and sustainability ($\beta = 0.261$) performance.

7. Discussion
The application of a lifecycle perspective within the research framework provides insights on the relative impact of each phase on diverse supplier performance. Previous survey-based studies separately took into account either design (Carr and Pearson, 1999; Cousins et al., 2008; Mahama, 2006; Heide et al., 2007) or implementation features (Prahinski and Benton, 2004; Prahinski and Fan, 2007). This study jointly includes the design, implementation, use and review of the SPMS. At a high level of analysis, the results suggest that the design, use and review affect supplier performance, thus confirming the need to effectively manage the entire SPMS lifecycle (Bourne et al., 2000; Gutierrez et al., 2015; Lohman et al., 2004; Braz et al., 2011). Nonetheless, going in depth within each single hypothesis, it is possible to gain interesting insights on the impact of each phase and to compare them.

The SPMS design, by far the most debated in extant literature, turns out to play a secondary role in respect to other phases, showing a significant positive impact only considering sustainability performance. Thus, H1 is only partially confirmed: in particular, only H1d is supported, while no significant evidence was found for H1a, H1b, H1c and H1e. What emerges is that a mature SPMS design per se does not improve supplier performance, no matter how much effort and emphasis is put into developing the right framework. As far as
Table III.
Measurement properties of the relative constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items asked of the respondent</th>
<th>Loading</th>
<th>Mean</th>
<th>SD</th>
<th>CR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPMS design maturity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence of operational metrics</td>
<td>We measure the extent to which the supplier delivers products/services according to buyer specifications</td>
<td>0.850</td>
<td>4.5</td>
<td>0.86</td>
<td>0.84</td>
<td>0.64</td>
</tr>
<tr>
<td>Presence of strategic metrics</td>
<td>We measure supplier flexibility performance</td>
<td>0.681</td>
<td>3.8</td>
<td>1.1</td>
<td>0.88</td>
<td>0.64</td>
</tr>
<tr>
<td>Maturity of the measurement process</td>
<td>Managers of other key functions actively participate in the design of the SPMS</td>
<td>0.791</td>
<td>3.9</td>
<td>0.78</td>
<td>0.87</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>SPMS implementation maturity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data collection reliability</td>
<td>The data collection process is rigorous and reliable</td>
<td>0.798</td>
<td>3.74</td>
<td>0.82</td>
<td>0.89</td>
<td>0.66</td>
</tr>
<tr>
<td>Frequency of reporting</td>
<td>We periodically report the SPMS data (either publicly or confidentially) to the supplier</td>
<td>0.885</td>
<td>3.37</td>
<td>1.1</td>
<td>0.9</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>SPMS use maturity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic use</td>
<td>We use the SPMS to monitor results</td>
<td>0.851</td>
<td>3.9</td>
<td>0.89</td>
<td>0.92</td>
<td>0.73</td>
</tr>
<tr>
<td>We use the SPMS to track progress towards goals</td>
<td></td>
<td>0.897</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We use the SPMS to compare outcomes to expectations</td>
<td></td>
<td>0.853</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive use</td>
<td>We use the SPMS to encourage discussion in meetings with the supplier</td>
<td>0.750</td>
<td>3.96</td>
<td>0.8</td>
<td>0.94</td>
<td>0.7</td>
</tr>
<tr>
<td>We use the SPMS to enable the organization and the supplier to focus on common issues</td>
<td></td>
<td>0.859</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We use the SPMS to enable the organization and the supplier to focus on critical success factors</td>
<td></td>
<td>0.825</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We use the SPMS to enable to launch continuous improvement plans</td>
<td></td>
<td>0.857</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We use the SPMS to develop a common language between our organization and the supplier</td>
<td></td>
<td>0.865</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We use the SPMS to develop a share strategy with the supplier</td>
<td></td>
<td>0.837</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SPMS review maturity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We periodically review the SPMS in view of the current competitive environment</td>
<td></td>
<td>0.891</td>
<td>3.2</td>
<td>1.01</td>
<td>0.94</td>
<td>0.83</td>
</tr>
<tr>
<td>We periodically review the performance measures to improve the clarity, usefulness and practicality of the SPMS</td>
<td></td>
<td>0.935</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We periodically review the SPMS to keep it aligned with the purchasing/SC strategy or corporate strategy</td>
<td></td>
<td>0.911</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
sustainability is concerned, buyer companies caring about this aspect generally measure the sustainability approach of suppliers by means of questionnaires and audits. They generally set some requirements that a potential supplier must respect and maintain over time.

Quite unexpected results emerge from the relationship between SPMS implementation maturity and supplier performance: a mature implementation does not have any significant impact on supplier performance, except for supplier innovation performance, where the effect is actually negative. It seems reasonable that the implementation of the system plays a minor effect on performance, if compared to system design, use and review. The negative impact on innovation performance may be explained as a supplier reaction toward a formal and frequent reporting (Prahinski and Fan, 2007), which could be perceived as excessively rigid, not favoring the innovation effort.

As far as SPMS use maturity is concerned, the related hypothesis (H3) is largely confirmed, finding a significant positive effect on supplier quality (H3a), delivery (H3b) and innovation (H3c). Coherent with the internal PMS literature, the combined diagnostic

<table>
<thead>
<tr>
<th>Construct</th>
<th>Items asked of the respondent</th>
<th>Loading</th>
<th>Mean</th>
<th>SD</th>
<th>CR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier quality performance</td>
<td>Our relationship with this supplier (buyer) has improved their (our) product quality</td>
<td>0.908</td>
<td>3.81</td>
<td>0.65</td>
<td>0.94</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Our relationship with this supplier (buyer) has improved their (our) product reliability and consistency</td>
<td>0.947</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Our relationship with this supplier (buyer) has lowered return rates on our (their) orders with them (us)</td>
<td>0.873</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplier delivery performance</td>
<td>Our relationship with this supplier (buyer) has improved on-time delivery of the orders we (they) place with them (us)</td>
<td>0.908</td>
<td>3.88</td>
<td>0.64</td>
<td>0.93</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Our relationship with this supplier (buyer) has improved delivery flexibility of the orders we (they) place with them (us)</td>
<td>0.885</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Our relationship with this supplier (buyer) has improved accuracy of delivery of the orders we (they) place with them (us)</td>
<td>0.908</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplier innovation performance</td>
<td>Our relationship with this supplier (buyer) has had a positive effect on his (our) ability to develop successful new products</td>
<td>0.937</td>
<td>3.52</td>
<td>0.7</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>Our relationship with this supplier (buyer) has had a positive effect on his (our) ability to make improvements to existing products</td>
<td>0.937</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplier sustainability performance</td>
<td>Our relationship with this supplier (buyer) has improved its (our) environmental sustainability performance</td>
<td>0.974</td>
<td>2.8</td>
<td>0.81</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Our relationship with this supplier (buyer) has improved its (our) social sustainability performance</td>
<td>0.975</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplier cost performance</td>
<td>Our relationship with this supplier (buyer) has provided us (them) with competitive prices</td>
<td>0.904</td>
<td>3.64</td>
<td>0.75</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Our relationship with this supplier (buyer) has reduced our (their) costs</td>
<td>0.924</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table III.
<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Presence of strategic metrics</td>
<td>0.801</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) SPMS interactive use</td>
<td>0.301</td>
<td>0.835</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Presence of operational metrics</td>
<td>0.138</td>
<td>0.438</td>
<td>0.798</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) SPMS diagnostic use</td>
<td>0.175</td>
<td>0.687</td>
<td>0.467</td>
<td>0.855</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) SPMS review maturity</td>
<td>0.283</td>
<td>0.520</td>
<td>0.323</td>
<td>0.563</td>
<td>0.912</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Supplier cost performance</td>
<td>0.208</td>
<td>0.269</td>
<td>0.083</td>
<td>0.154</td>
<td>0.234</td>
<td>0.914</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) Data collection reliability</td>
<td>0.181</td>
<td>0.452</td>
<td>0.331</td>
<td>0.616</td>
<td>0.520</td>
<td>0.107</td>
<td>0.814</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Supplier delivery performance</td>
<td>0.142</td>
<td>0.386</td>
<td>0.373</td>
<td>0.416</td>
<td>0.438</td>
<td>0.409</td>
<td>0.370</td>
<td>0.900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) Maturity of measurement process</td>
<td>0.370</td>
<td>0.534</td>
<td>0.309</td>
<td>0.547</td>
<td>0.437</td>
<td>0.116</td>
<td>0.496</td>
<td>0.144</td>
<td>0.836</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) Supplier innovation performance</td>
<td>0.212</td>
<td>0.260</td>
<td>0.151</td>
<td>0.197</td>
<td>0.278</td>
<td>0.348</td>
<td>0.015</td>
<td>0.466</td>
<td>0.081</td>
<td>0.937</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11) Supplier quality performance</td>
<td>0.214</td>
<td>0.367</td>
<td>0.268</td>
<td>0.360</td>
<td>0.334</td>
<td>0.330</td>
<td>0.224</td>
<td>0.592</td>
<td>0.162</td>
<td>0.681</td>
<td>0.910</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(12) Frequency of reporting</td>
<td>0.106</td>
<td>0.456</td>
<td>0.267</td>
<td>0.451</td>
<td>0.443</td>
<td>0.019</td>
<td>0.401</td>
<td>0.277</td>
<td>0.341</td>
<td>0.134</td>
<td>0.287</td>
<td>0.900</td>
<td></td>
</tr>
<tr>
<td>(13) Supplier sustainability performance</td>
<td>0.452</td>
<td>0.122</td>
<td>0.010</td>
<td>0.139</td>
<td>0.251</td>
<td>0.335</td>
<td>0.097</td>
<td>0.287</td>
<td>0.147</td>
<td>0.539</td>
<td>0.416</td>
<td>0.084</td>
<td>0.974</td>
</tr>
</tbody>
</table>

**Notes:** The square root of the AVE is shown in italics on the diagonal. Correlations are in the lower triangle of the matrix.
and interactive use maximizes the outcomes in terms of performance improvement: the interactive component allows improvement opportunity seeking in different areas; the diagnostic component strengthens the attention toward target achievement. When the two co-exist, there is positive pressure toward target achievement, which is shared by the buyer and the supplier with the goal of continuous improvement.

The SPMS review maturity plays a key role as well. $H_4$ is largely confirmed with a mature review showing a positive impact on supplier delivery ($H_{4b}$), innovation ($H_{4c}$) and sustainability ($H_{4d}$) performance, while no effect is registered on quality and cost performance ($H_{4a}$ and $H_{4e}$). The primary role of the review phase within SPMS lifecycle extends previous insights from a conceptual paper (Micheli and Manzoni, 2010) and case-based studies (Braz et al., 2011; Gutierrez et al., 2015). Reviewing metrics keeps the attention of the measured part high, avoiding any relaxing effect due to target achievement. The stimulus toward continuous improvement, along with constant attention toward the alignment with underlying goals, makes this phase so important.

The main result of the study is that the SPMS use and review (largely neglected in extant literature) have the largest effect on performance, together positively affecting supplier quality, delivery, innovation and sustainability performance. This seems coherent with ROT, in the sense that the orchestration role of the SPMS is particularly fulfilled within the use and the review phases. The former highlights the approach of the measuring part toward the measured counterpart, thus unraveling the ultimate purpose of the SPMS adoption in terms of the supplier's reaction. The latter enables the SPMS to match the inner

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Presence of strategic metrics</td>
<td>0.346</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) SPMS interactive use</td>
<td>0.211</td>
<td>0.761</td>
<td>0.594</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) SPMS review maturity</td>
<td>0.334</td>
<td>0.570</td>
<td>0.402</td>
<td>0.634</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Data collection reliability</td>
<td>0.261</td>
<td>0.313</td>
<td>0.112</td>
<td>0.179</td>
<td>0.274</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) Supplier delivery performance</td>
<td>0.225</td>
<td>0.515</td>
<td>0.445</td>
<td>0.719</td>
<td>0.692</td>
<td>0.131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Supplier diagnostic use</td>
<td>0.172</td>
<td>0.418</td>
<td>0.466</td>
<td>0.469</td>
<td>0.484</td>
<td>0.480</td>
<td>0.424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) Maturity of measurement process</td>
<td>0.457</td>
<td>0.625</td>
<td>0.414</td>
<td>0.661</td>
<td>0.321</td>
<td>0.168</td>
<td>0.615</td>
<td>0.163</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Supplier innovation performance</td>
<td>0.258</td>
<td>0.287</td>
<td>0.197</td>
<td>0.225</td>
<td>0.313</td>
<td>0.430</td>
<td>0.123</td>
<td>0.529</td>
<td>0.106</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) Supplier quality performance</td>
<td>0.252</td>
<td>0.402</td>
<td>0.334</td>
<td>0.406</td>
<td>0.369</td>
<td>0.290</td>
<td>0.259</td>
<td>0.061</td>
<td>0.194</td>
<td>0.775</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) Frequency of reporting</td>
<td>0.164</td>
<td>0.545</td>
<td>0.354</td>
<td>0.549</td>
<td>0.533</td>
<td>0.063</td>
<td>0.495</td>
<td>0.326</td>
<td>0.442</td>
<td>0.170</td>
<td>0.350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11) Supplier sustainability performance</td>
<td>0.513</td>
<td>0.135</td>
<td>0.072</td>
<td>0.151</td>
<td>0.271</td>
<td>0.384</td>
<td>0.162</td>
<td>0.317</td>
<td>0.172</td>
<td>0.598</td>
<td>0.454</td>
<td>0.100</td>
<td></td>
</tr>
</tbody>
</table>
dynamism of the orchestration process. The SPMS design and implementation often entail little or no interaction with the supplier; therefore the impact of these phases on supplier’s orchestration is less evident.

With respect to the previous findings, it is important to note that the maturity of the various phases could be inter-related. For example, a mature SPMS design may be positively related to a mature SPMS use or a mature implementation could positively impact the maturity of the SPMS review. On this behalf, it has to be noted that $R^2$ adjusted values for supplier performance dimensions (except for cost) are good and increase every time a lifecycle phase is added. Finally, supplier cost performance is not impacted by any of the SPMS lifecycle phases. An SPMS is put in place to manage the ongoing relationship with active suppliers. The suppliers’ price proposal and related savings are generally defined in the negotiation phase, in a pre-contractual situation.

8. Conclusions
The present study tests the existence of a positive relationship between SPMS adoption and supplier performance improvement, supporting the relevance of the SPMS in orchestrating supplier resources. Several contributions can be recognized as a part of the emergent literature stream regarding the performance impact of the SPMS. Indeed, both sides of the relationship under scrutiny have been investigated. On the buyer side, a lifecycle perspective has been applied to the SPMS, developing constructs for SPMS design, implementation, use and review maturity; on the supplier side, multiple performance dimensions (i.e. quality, delivery, innovation, sustainability and cost) have been considered separately. Along with a dyadic data collection, this allowed the achievement of more insightful and robust empirical evidence. We study buyer-supplier relationships through the lens of ROT, which provides a suitable theoretical framework to explain the role of a mature SPMS in orchestrating suppliers.

Important managerial implications are provided. Empirical evidence shows that the SPMS use and review have the widest impact on supplier performance. Strictly speaking, this means that measuring the right things is not enough; managers should instead actively use the SPMS (both for controlling and for stimulating collaboration) and timely review the metrics coherently with changing situations. For the buyer company, being able to properly manage multiple dyadic relationships with suppliers is of critical importance. Examining the entire SPMS lifecycle will support purchasing/SC managers in doing so.

The study displays some limitations as well, which open venues to further research on the topic. Regarding data analysis, the $R^2$ values of the model are in line with other studies, yet relatively low; thus, the importance of the significant linkages achieved must be relativized. The main aim of this study is to introduce a lifecycle perspective of SPMSs; therefore we test the direct link between the different lifecycle stages and supplier performance. The non-significance of some links and the $R^2$ values suggest that – in the broader context of buyer-supplier relationships – several factors other than a mature SPMS may affect performance. Here we can think about classical integration and/or collaboration practices as well as suppliers’ skills and any information technology tool supporting the interaction. In this sense, the SPMS might represent a necessary but not always sufficient condition to ensure supplier performance and the investigation should expand to include a broader set of factors (including boundary conditions) that can interact with the SPMS in the achievement of better results. Another direction for improving the level of significance can be the study replication and the extension/refinement of our SPMS lifecycle measures. To the best of our knowledge, this is the first study adopting a SPMS lifecycle perspective in the context of buyer-supplier relationships; therefore we see room for improving the scales and possibly identifying more significant aspects that can affect the supplier performance.
Since we mostly focused on the introduction of SPSM lifecycle measures and on the SPMS-performance link, we grounded on the buyer’s answers for the set of independent variables and on both buyer’s and supplier’s answers for the dependent variables, in order to ensure data triangulation and reliability. Therefore, the dyadic data are used for validation purposes rather than for the comparison of the different actors’ perception. Future studies could instead focus on the perception gap between the buyer and the supplier, and the causes/consequences of such gaps.

Also, the lifecycle perspective on the SPMS could benefit from an empirical investigation over time, to analyze how buyer companies manage supplier involvement all along the sequence of activities; therefore future studies could entail a longitudinal data collection process.

Finally, it could be interesting to look at the antecedents of a mature SPMS all along its lifecycle, thus understanding which resources and skills companies need to develop to properly measure and manage supplier performance.

References


Further reading

Corresponding author
Vieri Maestrini can be contacted at: vieri.maestrini@polimi.it

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Missing performance management and measurement aspects in performance-based contracting

A systematic process-based literature analysis of an astonishing research gap

Andreas H. Glas, Florian U. Henne and Michael Essig
Chair for Materials Management and Distribution, Bundeswehr University Munich, Neubiberg, Germany

Abstract

Purpose – Performance-based contracting (PBC) is a business model for the adaptive and innovative delivery of product-service systems. In PBC, the provider is paid according to the service performance with the aim of providing monetary incentives to safeguard possible outcomes as much as possible for the PBC customer. Performance measurement and its management are crucial for PBC success and, in particular, for the pay-for-performance link. However, the literature on PBC performance management is rather sparse, and there has been no systematic review on the topic. Thus, the purpose of this paper is to fill that gap and to present a comprehensive and systematic review of performance measurement and management in the PBC context.

Design/methodology/approach – The paper builds on a literature review based on a sample of 102 subject-relevant articles from academic journals. The content analysis follows a two-step procedure. First, the articles are coded following a process-based research framework. Second, the content of each process step is assessed in a qualitative text analysis.

Findings – The results show a surprising scarcity of papers that explicitly address performance management topics in the context of PBC. Only the topics of performance specification and performance indicators are broadly addressed, whereas in all of the other areas, e.g., strategic alignment, data capture and reporting, only limited specific findings could be found.

Research limitations/implications – The paper concludes that future research on performance management in PBC should expand its theoretical framework and empirical efforts in four specific proposed directions.

Originality/value – The paper provides an up-to-date review that is focused on performance management and measurement in the emerging context of PBC.

Keywords Literature review, Performance-based contracting, Performance measurement and management

Paper type Literature review

1. Introduction

Many companies in the business-to-consumer sector offer their clients access to and the usage of products but not ownership of such products (“use not own”). Typical examples are modern car sharing, couch surfing or music- or video-streaming offerings. Moreover, manufacturing companies in the industrial business-to-business sector offer more than physical products – they sell the performance of their products (Holmbom et al., 2014). The academic discourse addresses that phenomenon and discusses it under the term “performance-based contracting” (PBC) (Essig et al., 2016; Selviaridis and Norrman, 2014; Kim et al., 2007).

Interest in PBC has significantly increased in the last two decades (Selviaridis and Wynstra, 2015); however, the focus has been on clarifying the concept (Hypko et al., 2010) or analyzing its content, e.g., contract design (Datta and Roy, 2013). Although an increase in published PBC research papers has recently been observed (Selviaridis and Wynstra, 2015), the literature on PBC with a specific focus on performance measurement and management (PMM) is scarce, and there has been no systematic review that considers both aspects.
This situation is surprising with regard to two aspects. First, PBC explicitly addresses performance as a key element of its contract approach (Holmbom et al., 2014; Glas et al., 2013; Randall et al., 2011). It is important to have accurate PMM in place (Sols et al., 2007). There are numerous challenges to selecting, measuring and managing appropriate performance indicators (Holmbom et al., 2014; Tukker, 2004; Spring and Araujo, 2009). Typical key performance indicators (KPIs) in the context of PBC are, for example, availability, reliability, maintainability, supportability, logistics response time, logistic footprint and cost of use (Sols et al., 2007). Nevertheless, despite its high relevance, discussion of PMM is rare in PBC contributions. Furthermore, even the understanding of performance remains unsettled and the discussion of KPIs and their measurement is ongoing, recent reviews of the PBC literature have recommended placing PMM on the agenda for future research (Selviaridis and Wynstra, 2015; Mouzas, 2016).

Second, there is a need to further develop PMM given the dynamic change in environmental contexts and the emerging of new practices (Melnyk et al., 2014). PBC is not only considered to be a phenomenon of new and changed business practices but also an approach that represents the application of new and emerging theoretical methods such as service-dominant logic theory (Randall et al., 2010). In addition, contemporary supply settings make the PMM of suppliers and service providers essential for industrial customers (Gottschalk and Solli-Sæther, 2005). Nevertheless, the research on how source providers can set up and operate PMM is thus far generally lacking (Weimer and Seuring, 2009).

It thus appears that the research streams of PBC and PMM can profit from each other to further develop PMM in the context of service-oriented industrial business models and to apply state-of-the-art PMM knowledge in the emerging PBC field. Thus, the guiding research question is how the PMM literature can contribute to PBC and vice versa.

For this purpose, it is necessary to take stock of the existing literature, to assess how the field of PBC and PMM could effectively build on the existing literature and to identify which areas and research questions should be addressed. Therefore, this paper reviews and synthesizes the PMM literature within the PBC field. In particular, this research aims to expand knowledge on how deeply specific PMM topic areas are addressed within the PBC literature to identify the topics that represent research gaps. The next section provides a comparative review of the previous research on PBC and its relationship to PMM. Section 3 sets the stage for further analysis and reviews the PBC and PMM areas, building a contextual analysis framework for the review. In Section 4, the applied methodology is described. A review of the findings is presented in Section 5, and the findings are divided into quantitative and qualitative content analyses. In Sections 6 and 7, the findings are discussed, suggestions for future PMM and PBC research are proposed and managerial implications are outlined. Finally, Section 8 summarizes and offers concluding remarks on the study’s main contributions and limitations.

2. The previous research on PBC and its relationship to PMM

PBC and PMM appear to be closely related, and the relatively new concept of PBC can certainly profit from the existing research findings in the PMM literature. The field of PBC is an emerging topic for practitioners and academia, which explains the rising interest in and increase of publications on the topic of PBC (Selviaridis and Wynstra, 2015). The starting point of this research was an initial screening of the literature on PBC. On the one hand, the screening supported claims that aspects of PMM are rarely or only briefly touched upon in PBC contributions. On the other hand, the screening also identified eight literature reviews on PBC. To avoid repetitive research, this study briefly reports on these eight reviews and uses them to justify the applied methodology (Table I).
The oldest considered review is from Eldridge and Palmer (2009). The most recent reviews are from Alyami et al. (2015) and Selviaridis and Wynstra (2015). Although some reviews appear to be comprehensive, with more than 100 analyzed contributions (Holmbom et al., 2014; Selviaridis and Wynstra, 2015), other reviews only provide a narrow literature sample (e.g. 17 in the review of Eldridge and Palmer, 2009). Six reviews focus on one particular industry branch: two on construction, two on healthcare, one on aerospace in defense and one on manufacturing. Two reviews, which are partially based upon each other, focus on the functional aspects of operations and supply management (Selviaridis, 2011; Selviaridis and Wynstra, 2015).

The main objectives of the reviews are to define and classify PBC (Selviaridis and Wynstra, 2015; Hypko et al., 2010), to analyze benefits or drawbacks (Holmbom et al., 2014; Sultana et al., 2013; Selviaridis, 2011) and to report payment schemes (Carlson et al., 2010; Eldridge and Palmer, 2009). Only one review focused on performance measures (Alyami et al., 2015), and it reported highly specific findings, e.g., road roughness.

Overall, the previous content analysis in the context of PBC lacks a distinct focus on PMM. Even when aspects of PMM are reported (e.g. Alyami et al., 2015; Carlson et al., 2010),

<table>
<thead>
<tr>
<th>Authors</th>
<th>Journal</th>
<th>Industry focus</th>
<th>Main objective</th>
<th>Reviewed papers</th>
<th>Time scope</th>
<th>Relation to PMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alyami et al. (2015)</td>
<td>9th International Conference on Managing 2015</td>
<td>Construction sector</td>
<td>Review of PBC performance measures for road maintenance</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Very specific aspects are discussed (e.g. road roughness)</td>
</tr>
<tr>
<td>Holmbom et al. (2014)</td>
<td>Journal of Manufacturing Technology Management</td>
<td>Aerospace and defense</td>
<td>Review of benefits and drawbacks of PBL</td>
<td>101</td>
<td>1988-2013</td>
<td>PMM aspects are crucial implementation factors of PBC</td>
</tr>
<tr>
<td>Sultana et al. (2013)</td>
<td>International Journal of Productivity and Performance Management</td>
<td>Construction sector</td>
<td>Review on practices in PBC with focus on benefits and potential</td>
<td>62</td>
<td>n.a.</td>
<td>Very specific aspects are discussed (e.g. road roughness)</td>
</tr>
<tr>
<td>Selviaridis (2011)</td>
<td>Proceedings of the 20th Annual IPSERA Conference</td>
<td>No specific focus</td>
<td>General review within the operations and supply management research field</td>
<td>87</td>
<td>1985-2010</td>
<td>Investments in PMM within a PBC are seen as a research gap</td>
</tr>
<tr>
<td>Carlson et al. (2010)</td>
<td>Health Policy</td>
<td>Healthcare and welfare</td>
<td>Review of PBC schemes in healthcare</td>
<td>n.a.</td>
<td>1998-2009</td>
<td>Classification of schemes in terms of, e.g., health outcomes</td>
</tr>
</tbody>
</table>

Table I. Overview of previous literature reviews on PBC
the findings are industry specific, have little relationship to operations and supply management and thus are limited. The work of Selviaridis and Wynstra (2015) is the exception and addresses aspects of PMM in the supply management context; however, it does not review these aspects in detail. Instead, it uses them to conceptualize the authors’ understanding of PBC. They report in their appendix that 89 studies address or at least mention PMM aspects in their sample of 241 contributions (Selviaridis and Wynstra, 2015). Selviaridis and Wynstra (2015) is the starting point of this study because an in-depth analysis of PBC- and PMM-related literature has yet to be performed. This omission calls for a comprehensive review that links PBC with aspects of PMM.

3. Development of the analysis framework

This section describes the development of the framework that is used to further structure the topic and the review analysis. In recent years, a large number of academic papers have addressed PMM (Taticchi et al., 2010; Choong, 2014); thus, it is first necessary to discuss the understanding of performance and PMM. Then, the PMM process steps are presented as a basis for an analysis framework (Section 3.1). Next, the concept of PBC is defined and presented in more detail (Section 3.2). Finally, the analysis framework is developed (Section 3.3).

3.1 Review of PMM

3.1.1 Performance understanding. The term “performance” is not easy to define (Otley, 1999), and it has not been consistently defined in the management research (Krause, 2006; Lebas and Euske, 2007). The diverse character of performance definitions is summarized by Lebas and Euske (2007) who stated, “Performance per se may not be definable in the absolute. It is […] contextual both in terms of users and in terms of purpose.”

Specifically, performance is considered to be the degree of target achievement for the relevant stakeholders (Krause, 2006); however, performance is also about deploying and managing the components of the causal model(s) that lead to the timely attainment of stated objectives within constraints that are specific to the firm and to the situation (Lebas, 1995). Several authors define performance by referring to “efficiency and effective dimensions” and highlight the multidimensionality of performance (Lebas, 1995; Neely et al., 1995; Karrer, 2006; Lebas and Euske, 2007).

To obtain an understanding of performance in the context of PBC, in addition to the concepts of multidimensionality and the context dependency of performance, PBC’s relationship to processes must be included. Entchelmeier (2008) defined performance according to four process dimensions: input oriented, process oriented, output oriented and potential oriented. A similar definition of performance is described by Karrer (2006), who only replaced the “potential-oriented” dimension with an “outcome-oriented” one – in other words, the availability of a system or the number units that are produced or sold. The definition of performance as being output or outcome oriented will be of especial interest in the context of the analysis of PBC.

3.1.2 PMM. There is also no cohesivel view within the research community with regard to the definition of performance measurement (Choong, 2013). Tonchia and Quagini (2010) stated, “if we want to manage performance, we have to be able to measure it. […] That is, if you can’t measure it, you can’t manage it!” A definition is postulated by Franco-Santos et al. (2012) in which the emphasis is on the use of different performance measures and the link to strategy: “contemporary performance measurement comprises the use of financial and non-financial performance measures linked to the organization’s business strategy.” Neely et al. (2005) note that performance measurement is literally “the process of quantifying efficiency and effectiveness of action […].” As a summary of 27 definitions, Choong (2014) prescribed
Performance measurement “as meeting set targets (goals) with a view for improvement.” Performance measurement cannot be considered separately from performance management; they are interrelated and form an iterative process because an effective performance measurement system is designed around KPIs, considers variance analysis and cause-and-effect relationships and fosters decision making (Lebas, 1995; Choong, 2014). Following this integrated understanding, the term PMM is used for the remainder of the paper.

A comprehensive view of the content of PMM encompasses “formal and informal mechanisms, processes, systems, and networks used by organizations for conveying the key objectives and goals elicited by management, for assisting the strategic process and ongoing management through analysis, planning, measurement, control, rewarding, and broadly managing performance [...]” (Ferreira and Otley, 2009). Due to this content heterogeneity, PMM is often identified as a system to which several interdependent factors contribute. The main features of a PMM system are briefly described following Franco-Santos et al. (2007) and Choong (2013). These features are data (variables), measuring attributes and supporting infrastructure.

Data are often differentiated in quantitative and qualitative form in terms of their use for financial and non-financial purposes (Otley, 1999; Bourne et al., 2003). Neely et al. (1995) highlight the integration of different data into a “[…] set of metrics used to quantify both the efficiency and effectiveness of actions.” Measuring attributes are linked to strategic objectives and are used to generate the information that is necessary for decision making (Marshall et al., 1999). Many authors do not differentiate between measure, metric and indicator (Neely et al., 1995; Bourne et al., 2003; Choong, 2013) and use the terms interchangeably. The term supporting infrastructure incorporates not only the organizational setting and information systems but also methods of data collection and data analysis. In addition, a PMM system exhibits other characteristics. It must be easy to understand, retain a long-term character and be linked to a reward system (Gomes et al., 2004). Moreover, it must be flexible to adapt to strategic changes (Kennerley et al., 2003). These characteristics are used below in the discussion about PMM in the context of PBC.

3.1.3 PMM process. The process of PMM has been extensively analyzed in the literature (Kennerley et al., 2003). A simple PMM process consists of three phases: the design of measures, the implementation of measures and the use of measures (Bourne et al., 2000; Forslund and Jonsson, 2007). Neely et al. (1995) extended this perception and included not only the performance target setting but the need for feedback loops and a connection to reward functions. In line with this notion, Franco-Santos et al. (2007) identified five process categories, which include “selection and design of measures, collection and manipulation of data, information management, performance evaluation and rewards and system review.” In addition, Otley (1999) referred to the definition of strategic objectives and the assessment of strategy implementation, target setting, the design of an incentive system and finally the design of an informational system. An additional process step called the “refreshing” of the PMM system is justified by the organizational or environmental changes over time and their effect on the PMM design (Bourne et al., 2005). In this process step, a constant re-evaluation of the appropriateness of existing performance measures on the one hand and of the PMM system as a whole on the other is performed (Bourne et al., 2005).

The above-mentioned issues can also be used to assess a PMM process following the aggregated process view that is provided by Bourne et al. (2005) in which they distinguish seven main processes: linkage to strategic objectives, data capture, data analysis, interpretation of data, communication and information provision, decision making and taking action and refreshing.

Summarizing the literature on PMM and on the PMM process, this review defines performance as a multidimensional and context-dependent construct along input,
throughputs, outputs and outcomes. It also connects to the perspective that PMM is a system to which several factors contribute, such as measures, data or a supporting infrastructure. Finally, the literature on the PMM process is reviewed, and seven main process steps are identified that will be the basis for the analysis framework.

3.2 Review of PBC
This section provides a brief understanding of why there is a need for a new contract approach and how PBC is defined. It also provides initial indications that PMM is highly important for the execution of a PBC. Furthermore, a three-step contract process is presented.

First, academic interest in PBC has recently increased (Hypko et al., 2010; Sultana et al., 2013; Selviaridis and Wynstra, 2015; Nullmeier et al., 2016) because PBC addresses specific, contemporary challenges in buyer-supplier relationships and provides a solution mechanism for these challenges by aligning the goals of the supplier and the buyer by defining, measuring and rewarding outcomes (Selviaridis and Wynstra, 2015). However, the PBC literature remains heterogonous due to specific applications of the approach in various industries, for example, manufacturing (Hypko et al., 2010), defense (Glas et al., 2013), infrastructure (Schoenmaker and de Bruijn, 2016), energy (Hufen and de Bruijn, 2016) and transportation (Stanley and Hensher, 2008).

More specifically, the challenge is that customer companies no longer purchase products and services (separately) but rather seek integrated service bundles and products, which are often called “solutions” or “product-service systems” (Kleemann and Essig, 2013). The described shift from products to service bundles forces suppliers to continuously extend their service offerings and to develop toward a specific service provider. The contribution of suppliers and service providers is significant for the success of their customer companies and increases long-term dependencies (Hypko et al., 2010; Kleemann and Essig, 2013). This situation calls for new approaches to contracting (Baines et al., 2009).

In addressing this challenge, PBC plays a major role (Guajardo et al., 2012). The underlying idea of PBC is first that companies buy the outcome of a service solution and not a physical product with attached services (Datta and Roy, 2011; Randall et al., 2010; Kleemann and Essig, 2013). Second, the PBC provider in return is compensated (at least partially) based on a successfully achieved outcome (Doerr et al., 2005), for example, when an engine manufacturer is paid in relation to the usage hours, which depend upon engine uptime and reflect the customer value (Guajardo et al., 2012). Thus, a PBC provider is not tied to the detailed technical specifications of the customer; instead, the provider accepts full responsibility for how to achieve the desired outcome (Sols et al., 2007; Glas and Essig, 2008). This approach implies a transfer of risks to the PBC provider (Doerr et al., 2005) but, conversely, encourages the provider to improve performance during the contract period according to the implemented incentives and sanctions (Randall et al., 2014).

A PBC uses KPIs and incentives to improve outcome performance, which requires procedures to measure, report and manage the performance during the complete contract period (Datta and Roy, 2011). Thus, PMM is a vital aspect of each PBC and must be considered in all phases of the contracting process. However, there are many major challenges to PMM because poorly specified KPIs and incentives can lead to unintended consequences and foster provider opportunism (Kim et al., 2007; Neely, 2008; Selviaridis and Wynstra, 2015; McDonald and Roland, 2009). Moreover, a need for further research in the PMM area has been identified by various authors in the PBC field (Dean and Kiu, 2002; Weimer and Seuring, 2009; Sultana et al., 2013; Alyami et al., 2015; Selviaridis and Wynstra, 2015).

For the aims of this research, it is necessary to structure the PBC topic. PBC is basically a contracting approach. Therefore, the contracting process is used to divide PBC management into three phases: the design phase of PBC, the management and execution phase of PBC...
and the post-PBC phase (Lazzarotto et al., 2014; Selviaridis and Norma, 2015; Selviaridis and Wynstra, 2015). Each contracting phase can be linked to different PMM process steps, as is shown in the analysis framework (Section 3.3).

3.3 Analysis framework

To develop PMM in the emerging context of PBC, a deeper understanding of the relevant PMM aspects is required. There are existing frameworks for the analysis of PBC; however, they refer to risk management (Selviaridis and Wynstra, 2015), address the cooperation mode with supplier and life cycle costs (Glas et al., 2013) or solely focus on performance indicators (Sols et al., 2007). Thus, an analysis framework is developed in this section to structure and focus the content analysis of this study on all of the PMM aspects of a PBC.

For this purpose, the insights from PMM (Section 3.1) are merged with those from PBC (Section 3.2) into a process-oriented analysis framework. The framework is based on the seven steps of the PMM process (Bourne et al., 2005) but separates the design of the incentive system from the design of KPIs to highlight that important issue. Then, the three contract phases, design, management and post-contract (Lazzarotto et al., 2014), are connected with the eight PMM process steps.

Four PMM process steps are related to the design phase of the contract, which are as follows: (1) strategy alignment (2) performance specification (3) KPI design and performance targets and (4) design of the incentive system. The next three PMM process steps are then related to the management phase of the contract: (5) data capture (6) performance assessment and (7) performance reporting. The last PMM process step (8) taking action and refreshing, is of interest not only for the execution and management phase of a contract but also for the post-contract phase. Therefore, it is connected with both contract phases (Figure 1).

4. Methodology

4.1 Literature search strategy

The systematic literature review method is applied in this research (Denyer and Tranfield, 2009). A systematic literature review is a transparent process that ensures scientific rigor and replicability and outlines the audit process of the author's procedure and conclusions (Denyer and Tranfield, 2009). In preparing the review, the relevance and
scope of PMM within the PBC literature is determined through a scoping study (Tranfield et al., 2003). The objective of the scoping study was to identify the most relevant keywords that concern PBC. The papers that are relevant for the scoping study were selected based on the author’s knowledge of the field and an initial search within academic databases.

The identified keywords included “Performance-based contract*,” “Performance-based logistics,” “Outcome based contract*,” “Availability contract*” and “Performance-based payment.” Keywords such as “Servitization” or “Product service system” were not viewed as concentrating solely on PBC and thus were omitted. The review comprised articles from January 2000 to February 2017. That timeline comprises the most relevant articles that address PBC according to the previous reviews (Selviaridis and Wynstra, 2015).

The literature search employed the databases Emerald Insight, Science Direct and Ebsco Business Source Corporate Plus. The search filtered hits from academic, peer-reviewed journals. Additional papers were identified based on citations in the already found hits. Next, non-English language articles were excluded, which resulted in 482 articles in total. The abstracts of all 482 articles were read to ensure that all of the articles were actually related to the topic. Papers that concern PBC at an intra-firm level (e.g. management incentives) were eliminated. After discussions within the research team, all of the articles from journals on children or social work were excluded. These filters reduced the number of articles to 243. In a next step, the content of all 243 articles was reviewed, and articles that actually addressed PMM aspects were identified. This identification delivered the final sample for the literature review of 102 articles, which is in line with the previous findings of Selviaridis and Wynstra (2015), who found 89 hits on performance measurement within their sample of 241 PBC-related articles. Finally, the content of the articles was coded according to the analysis framework. The literature search strategy is summarized in Figure 2.

![Figure 2. Methodological approach](image-url)
4.2 Coding and content analysis

To cope with issues around the quality assessment of articles throughout the literature review, e.g., error or bias, the content of the articles was coded using standardized data extraction forms (recommended by Tranfield et al., 2003). General information (title, author and publication details), the methodologies, the industry focus of each article and content aspects that related to the analysis framework were coded. For the coding, a basic classification form was used with 8 classification categories and 32 classification objects in total. MAXQDA software and Excel coding tables were used for this task.

Generally, the authors followed the reliability and validity aspects of content analysis methodology (Krippendorff, 2012). As an example, the definition of each classification category was written into the classification form, which allowed each reviewer to understand the content of the coding in the same manner, thus increasing semantic validity. Additionally, the articles were reviewed by two researchers to prevent bias and enhance reliability. In a first step, 20 articles were coded independently by both researchers according to 32 classification objects. Agreement on 29 classification objects between the authors was reached, which indicated an inter-rater reliability of 90.6 percent, which is an acceptable percentage for further coding (Carey et al., 1996). All of the disagreements were discussed to clarify the interpretation of the categories. The remaining articles were divided between the researchers to finalize the coding. This activity also included regular meetings to discuss research findings, emerging issues on the application of the categories and the general progress of the research.

5. Findings

5.1 Sample descriptive

The literature search revealed 102 papers that address PMM in the PBC context. The distribution of the papers over time confirms the previous findings and shows a recent steep increase of published scientific papers (Selviaridis and Wynstra, 2015). The topic has clearly been increasing in academic relevance since the mid-2000s, and the hits in 2015 and 2016 indicate that this trend continues (Figure 3).

Next, the findings suggest a dominance of qualitative and conceptual approaches (Figure 4). A total of 62 papers (61 percent) report on such methods, while quantitative (18 percent) and mathematical (14 percent) contributions are underrepresented. This literature review also includes eight papers (8 percent) that conducted their own reviews. As mentioned above, these reviews clarify the PBC concept (Hypko et al., 2010; Holmbom et al., 2014; Selviaridis and Wynstra, 2015) or address sector-specific issues (Sultana et al., 2013; Alyami et al., 2015). No specific and profound review on PMM in the context of PBC was found.
Sector-specific issues are often discussed in the literature, e.g., PBC in the defense context is called performance-based logistics (Glas et al., 2013). The articles cover diverse industry sectors (Figure 5). Aerospace, defense, healthcare and construction are often referred to, and energy, transportation, manufacturing and the public sectors are of high relevance. The housing and the oil and gas industries play a minor role. A total of 16 papers (16 percent) discuss PBC across industries or without explicitly referring to one specific industry.

However, the findings show that the literature on PMM in the PBC context is not consolidated (Figure 6). The 102 hits are distributed across 64 different journals. The journals address different disciplines, topics and target groups, such as logistics (Transportation Research Record), manufacturing (Journal of Manufacturing Technology Management), production (Journal of Production Management) and a few PMM outlets (International Journal of Productivity and Performance Management). Only in 22 journals were two or more articles published, which is also an indicator of the scattered field.

5.2 Content findings
The analysis reveals the status quo of academic attention on specific PMM process steps. Overall, the 102 papers address 257 content issues that are related to the PMM process. The topic of performance specification (step 2) and KPI design and performance targets...
(step 3) are the areas of highest interest, with 64 and 60 codings, respectively, which represent the number of references to these topics, and they are by far the highest number of references. Additionally, four topics (design of incentive system, data capture, performance assessment and taking action and refreshing – steps 4, 5, 6 and 8) are discussed with a number of codings (21 to 33 each). These topics are addressed, even when most of the contributions only touch on the topic. In contrast, the topics of strategy alignment (step 1) and performance reporting (step 7) are addressed by only 13 and 9 codings, respectively. Additionally, these topics are only addressed in side notes (Figure 7). The distribution of content hits by industry branch indicates no significant distribution differences across industries, even when the aerospace/defense and healthcare/welfare industries dominate the discussion by numbers, at least for the PMM process steps 2 – performance specification, 3 – KPI design and performance targets and 4 – design of incentive system (see Table Al).

5.2.1 Strategy alignment. An essential element in the PMM process is the definition of the strategic objectives of the company as a first step (Otley, 1999). Most authors who address PBC do not stress alignment with corporate objectives at all; only 13 codings mention that aspect. However, some authors argue that an effective PBC is aligned to the strategic goals of the customer (Buchanan and Klingner, 2007). The same understanding of the linkage of incentives with corporate goals is described in a case study by Selviaridis and Norrman (2014): “This aligns […] PBC provider’s incentives to customer’s strategic goal of making products available at the lowest possible cost.” There is a need to align the performance and the respective KPIs to the corporate goals of the buying organization (Barber and Parsons, 2009). However, it is often a challenge in the design of a PBC to
actually link KPIs with corporate goals (Selviaridis and Wynstra, 2015). In practice, KPIs often do not represent the corporate goals of the buying organization (Datta and Roy, 2011; Selviaridis and Norrman, 2015).

The linkage of KPIs, incentives or even the whole contract to strategic targets is also mentioned in the public or transportation sector. In these industries, the PBC must be aligned with the government objectives, as is stated by Stanley and Hensher (2008):

“...maintaining performance pressure for cost effective service delivery.” Particularly in the public sector, authors also emphasize the importance of linking the public purpose in general to the performance specification of the contract (Sols et al., 2007; Stanley and Hensher, 2008).

Overall, only 13 codings address the issue of strategy alignment. There is broad consensus that, in a PBC context, the incentives and KPIs should be aligned with strategic goals. However, there is no further information in the literature with regard to how the alignment of a PBC and strategy goals can be ensured.

5.2.2 Performance specification. With 64 codings, the aspect of performance specification appears to be crucial for PBC. A broad definition of performance is applied to PBC because performance is understood as a form of outcome or output, of efficiency and effectiveness and as a multidimensional construct that consists of several areas of performance (Sols et al., 2007; Selviaridis and Normann, 2014; Lu, 2016). Many authors within the field of PBC describe the focus of the customer as a performance goal that is defined as an outcome or as his expectation of the performance to be delivered (Kim et al., 2007; Hynk et al., 2010; Randall et al., 2011; Glas et al., 2013). However, in the majority of cases, a definition of the term outcome is missing (Martin, 2007; Hynk et al., 2010; Deng et al., 2014; Essig et al., 2016).

For example, Vitasek and Manrodt (2012) only state, “[...] a service provider is paid only when it is successful in achieving the mutually agreed desired outcomes.” Moreover, Caldwell and Howard (2014) describe performance as only “the outcomes required from the contract.” However, other authors try to describe the term performance. Selviaridis and Normann (2014) describe performance as an output or outcome; output refers to “the service functionality and level of performance (e.g. machine availability percent)” and outcome refers to “customer value derived from a given service.” In contrast to this definition, Glas et al. (2013) presented a broader definition of outcome. They defined an outcome as “[...] either process-related...
(e.g. service time), potential-related (e.g. availability) or result-related (e.g. operating time or mission success). In this context, other authors also subsume the level of performance, for example the availability of a system, to the definition of outcome (Caldwell and Howard, 2014; Randall et al., 2014). Others do not distinguish between outputs and outcomes at all and mention both terms within the same context (Hypko et al., 2010; Sultana et al., 2013).

Another aspect of performance is the importance of efficiency and effectiveness because in a PBC context, both must be ensured (Dean and Kiu, 2002). In the defense sector, the military aspect requires that the system be effective, and this includes characteristics such as robustness and resilience as well as efficiency in delivery (Glas et al., 2013). However, in the energy sector, performance is always defined as making investments in energy efficiency for the customer, which leads to overall improved plant and operational efficiency, and obtaining income from a proportion of the saved energy costs (Davies and Chan, 2001; Deng et al., 2014).

Finally, performance is also considered from a multidimensional perspective; that is, it is necessary to consider multiple factors such as quality, cost, delivery and timeliness to evaluate supplier performance (Buchanan and Klingner, 2007; Lazzarotto et al., 2014). Lu (2016) argues that it is difficult to find suitable measures that incorporate the full spectrum of human service performance, which is characterized by multidimensional and competing values. The outcome in general can be defined by several, individually determined, dimensions, for example, operational availability, logistics response time and customer satisfaction (Doerr et al., 2005; Sols et al., 2007).

In summary, the review showed that there is confusion about the exact definition of performance within PBC that is also dependent upon the industry sector. The terms “output” and “outcome” are occasionally used interchangeably, or at least they are not defined. As emphasized by the PMM literature, the authors within the PBC field are aware of the multidimensionality of performance and highlight the efficiency and effectiveness of the demanded outcomes.

5.2.3 KPI design and performance targets. The above-noted aspect of PMM is also of high importance for PBC due to the relatively high number of 60 codings. In the design phase of PBC, the objective is to encourage the contractor to optimally deliver the requested performance (Stanley and Hensher, 2008). In this respect, the KPIs that represent the requested performance (desired outcome) are a vital element to measure the achieved performance in relation to predefined performance targets (Glas et al., 2013; Alyami et al., 2015). The literature states that the selection of KPIs that actually represent the desired outcome is a major step in the design and management of PBC (Datta and Roy, 2011). KPI definition often represents a key challenge for the customer (de la Garza et al., 2009; Caldwell and Howard, 2014; Holmbom et al., 2014). For example, when system reliability is measured using the KPI “mean time between failures” (MTBF), it is often difficult to define what a failure actually is (Holmbom et al., 2014). To analyze the MTBF, other low-level KPIs such as failure diagnostics time or parts delivery time must be considered (Jin and Tian, 2012).

It is possible to assess contract performance using a single KPI, which is often the case in the utilities sector (Sols et al., 2008). However, in most cases, performance is measured across industries by a multidimensional set of KPIs that can incorporate quantitative and qualitative elements (McLellan et al., 2008; Datta and Roy, 2011) and can be objectively or subjectively measured (Holmbom et al., 2014). However, this set of KPIs should be limited to a few high-level KPIs that form the basis for incentives (Vitasek and Manrodt, 2012; Kristiansen, 2017).

With regard to the design of specific KPIs for PBC, several requirements have been identified in the literature (Alyami et al., 2015; Selviaridis and Wynstra, 2015): a KPI must be clearly defined and documented, outcome oriented, linked to customer strategic objectives,
SMART (specific, measurable, achievable, result-oriented and timely), objectively measurable and the number of KPIs in use should be limited. Furthermore, numerous articles report on PBC and the KPIs that are used in the practice, often based on case studies. According to Selviaridis and Norman (2015), common KPIs in the service supply chain are packaging delivery precision percentage, packaging availability percentage, transport delivery precision percentage, cost savings for customers, transport delivery accuracy percentage, picking accuracy percentage and perfect order percentage. The KPIs of a logistics service provider for a military system are availability, reliability, maintainability, supportability, logistics response time, logistic footprint and cost of use (Sols et al., 2007; Nowicki et al., 2010). Cleaning services in public transport are assessed by the following KPIs: quality (cleanliness of trains), safety (processes and tools/materials used) and personnel (employee satisfaction) (Nullmeier et al., 2016). The industry perspective indicates that multiple performance indicators are required in a PBC setting (Sols et al., 2007; Nowicki et al., 2008).

The KPIs were grouped to structure the content of the reviewed articles. Some articles only report on generic KPIs, e.g., “availability” or “mobility,” whereas other articles explicitly mention specific KPIs, e.g., “shipments volume per pallet” or “MTBF.” Therefore, the findings are presented in generic KPI categories (Table II) and specific KPI measuring specifications (Table III).

<table>
<thead>
<tr>
<th>Focus area</th>
<th>Key performance indicator categories</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistics in defense and aerospace</td>
<td>Availability, reliability, maintainability, supportability, logistics response time, logistic footprint and cost of use</td>
<td>Sols et al. (2007), Barber and Parsons (2009), Nowicki et al. (2010)</td>
</tr>
<tr>
<td>Logistics in defense and aerospace</td>
<td>Utilization, productivity, effectiveness, transactions</td>
<td>Doerr et al. (2005)</td>
</tr>
<tr>
<td>Logistics in defense and aerospace</td>
<td>Second-tier metrics for system effectiveness, operational availability, availability, degraded availability, mission availability and degraded mission availability, operational reliability, mission reliability, mission dependability, degraded mission reliability and degraded mission dependability, logistics footprint; response time</td>
<td>Sols et al. (2008)</td>
</tr>
<tr>
<td>Construction works</td>
<td>Mobility (traffic flow), safety (residents, construction workers), accessibility (duration of construction on site) and environment (noise, dust, waste, reused materials, total use of materials)</td>
<td>Sebastian et al. (2013)</td>
</tr>
<tr>
<td>Train manufacturing</td>
<td>Operational availability (trains are to be ready for departure at specified times), mission reliability (the trains should arrive at their destination stations on time, provided they departed from the origin stations on time)</td>
<td>Sols et al. (2007)</td>
</tr>
<tr>
<td>Industrial maintenance</td>
<td>Quality performance, cost, delivery time</td>
<td>Lazzarotto et al. (2014)</td>
</tr>
<tr>
<td>Cleaning services in transportation</td>
<td>Quality (cleanliness of trains), safety (processes and tools/materials used), personnel (education provided and employee satisfaction)</td>
<td>Nullmeier et al. (2016)</td>
</tr>
<tr>
<td>Road maintenance</td>
<td>Level-of-service effectiveness, cost efficiency, timeliness of response (response time of the contractor to service requests), response to complaints, response to emergencies and snow removal, safety procedures, quality of services (assesses the customer perceptions with respect to the condition of the assets and contractor performance)</td>
<td>de la Garza et al. (2009), Ozek et al. (2010), Sultana et al. (2013)</td>
</tr>
<tr>
<td>Road maintenance (pavement)</td>
<td>Pavement evaluation indices, e.g., pavement condition index (PCI), mathematical equations of which the inputs are values of different performance measures such as rutting and cracking</td>
<td>Alyami et al. (2015)</td>
</tr>
</tbody>
</table>

Table II. KPI categories in PBC
Once the relevant KPIs are selected and defined, baseline values and required performance targets must be established (Sols et al., 2008; de la Garza et al., 2009). Baselining each KPI could already be a challenging task in a PBC. The baseline can be based on historical performance data; however, even that might not be reliable when the measurement of the data has been done differently or the system is now used differently (Sols et al., 2008; Holmbom et al., 2014). For newly developed systems or service bundles with no historic experience, baselining is even more challenging because the baseline must be defined based on predictions of future performance (Kim et al., 2007).

After setting the baseline – the “normal” performance of a system – the next step is to define the desired performance target (Holmbom et al., 2014). This target is connected with financial incentives in a PBC and is often a topic of negotiations and discussions between the PBC provider and the contractor, particularly when the performance targets have been set too high (Selviaridis and Norrman, 2015). However, setting the performance targets too low is also risky because both sides are satisfied by good reported performance, but no one challenges the indicators in terms of whether an even better
As Holmbom et al. (2014) stated, “Performance target values must represent the customer's needs and form a realistic challenge for the supplier.” In addition to the described performance targets, defining and setting minimum performance levels below which the customer is allowed to drastically penalize the contractor or even terminate the contract is also recommended (Sols et al., 2008).

Several sources address the issue of how to implement PBC as a business model (e.g. Kumar and Markeset, 2007; Jin and Wang, 2012; Freund and Stölzle, 2016); however, few sources address the more detailed aspects of the implementation of a PBC KPI measurement system. There are empirical findings, e.g., the need for investments (Selviaridis and Normann, 2014); however, the topic of KPI implementation in PBC is clearly a gap in the literature.

Overall, KPI design and performance targets are broadly discussed in the PBC literature, compared with other PMM process steps. There are different opinions with regard to how to measure performance (single KPI vs several KPIs); however, the literature broadly agrees on KPI requirements and the procedure to define KPI baseline or target values. A multitude of KPIs are mentioned in the literature, whereas there is only limited discussion of implementation aspects.

5.2.4 Design of an incentive system. The design of a PBC payment system determines the degree of incentive alignment by defining the extent and intensity of bonuses and penalties (Kim et al., 2007; Sols et al., 2007). The objective is to balance several KPIs and transform the results into a reasonable payment (Sols et al., 2008), although it is a challenge to find a performance incentive with the necessary but reasonable intensity (Selviaridis and Normann, 2015). A PBC incentive system requires a financial performance payment component (Doerr et al., 2006; Glas et al., 2013). This financial incentive is linked to KPIs to motivate the contractor to perform in the interest of the customer (Hünerberg and Hüttmann, 2003; Selviaridis, 2016). The extent of the variable component of the payment and incentive system is highly project specific. A payment that is directly linked to KPIs is in fact conceivable; however, in practice, partially tied payment systems dominate (McLellani et al., 2008; Glas et al., 2013; Selviaridis and Wynstra, 2015).

For the manufacturing industry, Hypko et al. (2010) distinguish the pay-on-availability (provided performance), pay-on-production (demanded performance) and pay-per-use (demanded performance) PBC payment approaches. The pay-on-availability approach ensures the customer the availability of a specified system or machinery. Following the pay-on-production approach, contractor compensation is tied to the manufactured output, whereas the pay-per-use approach focuses on how often a system is used or on the volume of manufactured products (Hypko et al., 2010).

There are essentially five payment models that are described in the PBC literature (Figure 8): fixed-price incentive fee (FPIF), cost-plus incentive fee (CPIF), firm-fixed-price (FFP), exclusively performance-based (EPB) and based on economic result (BER).

The most common payment model is the FPIF, which provides a fixed price for a defined performance (e.g. system availability) based on a defined level of use (e.g. hours in use) and adds an incentive component when performance targets are achieved or surpassed (Kim et al., 2007; Nowicki et al., 2010; Caldwell and Howard, 2014).

A CPIF is determined by reimbursed costs for the service with an additional incentive component to foster performance improvements or cost reductions (Nowicki et al., 2010; Datta and Roy, 2011; Glas et al., 2013). These types of contracts can also be used as a first step to a FPIF payment model (Kim et al., 2007).

Some authors actually do not consider a FFP to be a PBC payment model (Kim et al., 2007; Hypko et al., 2010); however, when the fixed price is tied to minimum performance targets
that must be achieved, it can also be considered to be an alternative PBC payment model (Randall et al., 2011; Glas et al., 2013).

To integrate the KPIs with the described payment models, Sols et al. (2007) suggested the concept of penalty, dead and reward zones. Only when the performance reaches the reward zone does the contractor realize a superior profit. In the dead zone, he/she is reimbursed according to the agreed price, and in the penalty zone, the price is cut (Sols et al., 2007).

In addition to the described payment models, which are partially tied to performance, there are also payment models that are discussed that fully tie performance to payment (EPB). This model best fits the core concept of PBC because it focuses the contractor solely on achieving the defined performance. An EPB is often neglected by contractors due to the high-risk transfer (McLellan et al., 2008; Glas et al., 2013).

The last discussed model in the PBC literature is a payment model in which the fee is based on customer economic result (BER). In this model, performance is tied to defined KPIs that reflect the customer economic result, for example generated revenues, saved costs or contribution to customer margin (Hünerberg and Hüttmann, 2003; Hypko et al., 2010).

In summary, the analyzed literature essentially agrees that the design of an incentive system is important for PBC. However, several payment schemes are presented, and the schemes are named differently (e.g. pay-per-use vs incentive fee payment). There is also no consensus with regard to which payment schemes are actually PBC. Overall, the PMM topic is not yet consolidated for PBC.

5.2.5 Data capture. The next step in the PMM process is to check whether and how the data for a KPI can be measured and whether the data are precise, accurate and reliable (Choong, 2014). Such issues are often an obstacle in designing an effective PMM system in a PBC context (Selviaridis and Normann, 2015). The review shows that in most of the articles, data capture does not play a role because it is hardly mentioned. Only 26 codings (out of 257) mention the topic of data capture.

In these articles, several means of data capture are described. Nullmeier et al. (2016) identified two main approaches to gather contractor performance data, including contractor audits and customer satisfaction surveys; however, they acknowledge that additional methods are occasionally used to verify the same performance category to increase credibility for the contractor. In road maintenance or construction, the data are primarily gathered through field inspections (de la Garza et al., 2009; Mouzas, 2016). In addition to audits, surveys and field inspections, conducting meetings to gather data directly from the responsible persons is also suggested (Kumar et al., 2006); however, the credibility of the data with regard to the influence on incentives is questionable. If performance cannot be objectively verified by auditors, contractors might encounter audit exceptions, questioned costs and demands for
repayment (Martin, 2007). Conversely, the customer then must ensure that the PBC contractor is not gaming the system, e.g., when a bus driver does not stop at all of the bus stops because he/she is late, which is directly related to financial bonuses (Greiling, 2006).

Another interesting question is related to the responsible party performing the data gathering. It might be assumed that in a PBC, the data gathering is performed by the customer due to the link with payment; however, this approach is not always used (Alyami et al., 2015; Nullmeier et al., 2016). In addition, contractors can be held responsible for periodically collecting performance data (Martin, 2007). Then, the reliability of the data gathering must be ensured, e.g., by randomly joining the contractor during data gathering or performing additional quality audits (Alyami et al., 2015). Another approach is third-party data collection, which is often assumed to be more objective; however, additional costs are incurred with that approach (Stenbeck, 2008).

When measuring performance, the most important element is the underlying IT infrastructure, which can determine a reliable, accurate, effective and efficient data collection process (Randall et al., 2010; Selviaridis and Norrman, 2015; Mouzas, 2016). Particularly in PBC contexts that include complex and expensive product-service bundles, real-time data collection that is supported by IT is necessary (Forslund, 2012). Moreover, in the field of energy service contracts, high investment in IT infrastructure for data gathering and auditing is compulsory, often requiring specialized tools and equipment (Sorrell, 2007). In this context, the additional investments in data collection and measurement systems can outweigh the benefits that can be gained by extra rewards and must be thoroughly considered (Greiling, 2006; Kumar et al., 2006).

Overall, it is surprising that data capture is hardly mentioned in the reviewed literature because this PMM process step is considered to be a possible obstacle to designing effective PBC. The literature sources agree that there are multiple approaches to gathering data, e.g., audits, surveys, field inspections or systems that are fully automated by IT. There is no clear consensus or recommendation with regard to data gathering responsibilities. Discussed alternatives include data gathering that is performed by the customer, by the provider or by a third party. Furthermore, studies broadly agree that an underlying IT infrastructure is helpful for PMM in PBC. However, the names of the IT systems that are in use are not mentioned, nor are the requirements or functionalities of such systems discussed. Altogether, the literature on data capture is consistent on general aspects (e.g., the need for IT infrastructure and the need for the clarification of responsibilities); however, further guidance (e.g., evaluation or even case descriptions of PMM IT systems) remains missing.

5.2.6 Performance assessment. After the performance data are gathered, they must be analyzed to provide feedback to the contractor and report results to respective stakeholders (Sols et al., 2008; Molenaar and Navarro, 2011). This PMM process step has received surprisingly little notice (21 codings), even when it is necessary to develop and implement monitoring methodologies, procedures and tools (Alyami et al., 2015) that are structured systematically (Kumar et al., 2006). Performance monitoring and assessment procedures are aimed at the evaluation and definition of service success, assigning failures to specific actions of the contractor or the customer and investigating whether the contractor is accountable for these failures (Selviaridis and Norrman, 2014). In regular assessment meetings, the achievement of performance targets and deviations from targets are discussed, and corrective actions are jointly planned (Kumar et al., 2006; Buchanan and Klingner, 2007). In preparation for these meetings, data assessment can be structured in three phases (Larbi, 2001; Sols et al., 2008; de la Garza et al., 2009), which are as follows:

1. calculation of actual and target values of the respective KPIs;
2. performing a comparison of actual values and target performance values; and
performing a comparison between different areas of service, different years (also trend analysis) and in relation to budget with the objective of deriving further actions.

The aim of these steps is to transform the gathered data into information (Sols et al., 2008). The assessment of the information in meetings is of high importance, and it is considered to be a critical success factor (Selviaridis and Normann, 2015). Overall, performance assessment is a necessary step within the PMM process, and the extent of data analysis must be considered in the context of the required investments of time, effort and money (Ozbek et al., 2010). Some sources provide further insights into performance assessment, e.g., the execution of performance assessment meetings.

5.2.7 Performance reporting. Overall, only nine codings address performance reporting. Some of them explicitly address the issue and identify several aspects (Greiling, 2006; de la Garza et al., 2009) while other sources only mention some aspects. First, several authors identify the need to report performance (Greiling, 2006). Selviaridis and Wynstra (2015) and Molenhaar and Navarro (2011) measure and report performance; however, only Greiling (2006) notes the high importance of reporting as an “important driver for the introduction of performance measurement.” Greiling (2006) also weighs the efficiency and effectiveness of reporting. Ineffectiveness is mentioned, for example, the production of reports that do not influence the decision-making process or increase efficiency generate high costs for gathering information that outweigh the benefits of the data.

Another aspect is reporting frequency. In the articles, there are four comments on the frequency of performance reporting: Larbi (2001), Vitasek and Manrodt (2012), Selviaridis and Wynstra (2015) and Kumar et al. (2006) simply state that reporting should occur regularly. However, it is not clear what “regular reporting” actually means; a whole range, from real-time to daily, monthly, quarterly or annual reporting, is conceivable.

Additional information is given with respect to the design and/or formatting of reporting. de la Garza et al. (2009) distinguish three alternative report designs: report card with a simple presentation of KPIs; report of effectiveness, with a detailed analysis of the PBC performance; and report of deficiencies with critical issues addressed to the PBC provider. Ozbek et al. (2010) also refer to the report card as a meaningful reporting design and provide deeper insights into the reporting formatting by proposing the use of as many visuals as possible. Reporting should produce appropriate and user-friendly reports that can effectively communicate the results to stakeholders (Ozbek et al., 2010).

Finally, there are several comments on the content of performance reporting in PBC. Greiling (2006) indicates that the reporting of multidimensional KPIs would enhance transparency. de la Garza et al. (2009) note the reporting of KPIs on effectiveness and on performance deficiencies.

Overall, this section showed that the academic discussion of the performance reporting of PBC remains in its infancy because the included aspects of reporting (need for reporting, efficiency and effectiveness, frequency, organization, design and/or formatting and content) are only slightly touched upon. Most statements are based on generic or plausible clues that concern how to report management content. A broader discussion on how performance should be reported in PBC is missing.

5.2.8 Taking action and refreshing. Issues that concern the PMM process, step-taking action and refreshing are limited to the following aspects. First, the issue of contract extension or termination is supposed to be relevant (Lazzarotto et al., 2014). In PBC, the opportunity for an early contract extension can provide additional incentives for the contractor (Geary et al., 2010). This type of incentive can also raise the issue that once the contract extension is granted, it loses its power. To cope with that problem, multiple contract extensions are suggested (Schoenmaker and de Bruijn, 2016). Conversely, an early
contract termination can also be considered when the contractor repeatedly misses the minimum performance targets, dependent on how this alternative is defined in the contract (Sols et al., 2008).

A challenge of long-term PBC is to develop a PMM and incentive system that remains relevant, achievable and challenging over the entire contract period (Geary et al., 2010; Hufen and de Bruijn, 2016). Therefore, several authors emphasize the importance of flexible contract configuration and the corresponding PMM system (Kumar et al., 2006; Sols et al., 2007; Datta and Roy, 2011; Vitasek and Manrodt, 2012). Selviaridis and Wynstra (2015) discovered that “contractual flexibility is required to adjust performance measures and financial incentives as experience accumulates.” Datta and Roy (2011), Selviaridis and Normman (2014) and Selviaridis (2016) found that a certain flexibility of the defined performance targets results in increased system performance for the customer and general improved performance of the contractor, which leads to a win-win business relationship. Moreover, the contractor is more amenable to risk taking (Selviaridis and Normman, 2014). In addition to these internal factors, there could be changes in the environment that influence the contract (Randall et al., 2010; Batista et al., 2016). In this context, Sols et al. (2007) argued, “[…] it is important to ensure that a system’s supportability strategy and corresponding performance-based logistics methodology are robust and agile, easily adapting to the evolving operational environments.” Due to the individual specification of each PBC and the often-missing experience at the beginning of the contractual relationship, decreasing the contract flexibility is recommended as the contract duration progresses (Hufen and de Bruijn, 2016). Although many authors stress the need for flexible contracts, an increase in contract complexity is reported (Schoenmaker and de Bruijn, 2016), which in turn underscores the importance of PMM.

In summary, the PBC literature only focuses on certain actions with regard to contract flexibility and on discussions of an early contract extension as an additional incentive instrument. These issues are identified as possible challenges for the development of PBC contracts over time. An in-depth analysis of solution mechanisms for these challenges is missing in the analyzed literature.

6. Discussion
The starting point of this work was the initial suspicion that the research streams of PBC and PMM are barely linked. From a quantitative point of view, this review identified 102 contributions and 257 content coding references that address PMM issues within PBC-focused articles. The sheer number might indicate that there is a broad basis of literature that connects PMM with PBC. However, the content analysis revealed that the maturity level of the PMM discussion in the context of PBC varies greatly depending upon the PMM process step.

The PMM process steps 2 – performance specification, 3 – KPI design and 4 – design of the incentive system are discussed more broadly within the PBC articles. However, these process steps only have a loose relationship with the specialized PMM literature. References to common PMM frameworks such as the balanced scorecard (Kaplan and Norton, 1992) or the performance pyramid (Cross and Lynch, 1988) cannot be reported at all within the analyzed literature. The other PMM process steps are not substantially discussed. These topics are often mentioned without any conscious reference to the PMM literature. Overall, the content findings of this review generally confirm the initial suspicion and indicate that there is only a minimal connection between the literature streams of PBC and PMM (see Table IV).

More specifically, the findings from the content analysis provide answers for the study’s guiding research questions with regard to how the PMM literature can contribute to PBC.
<table>
<thead>
<tr>
<th>PMM process step</th>
<th>PBC phase</th>
<th>Typical statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Strategy alignment</td>
<td>Contract design</td>
<td>“A potential specification challenge is a gap between the strategic mission and aims of the buying firm and the performance metrics” (Selviaridis and Wynstra, 2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“An effective performance contract motivates contractors to help achieve the owner’s business goals and objectives through financial incentives” (Buchanan and Klingner, 2007)</td>
</tr>
<tr>
<td>2. Performance</td>
<td></td>
<td>“[…] the customer has to consider monitoring the performance of contracted services to assure the effectiveness and efficiency of the outcome” (Dean and Kiu, 2002)</td>
</tr>
<tr>
<td>specification</td>
<td></td>
<td>“PBL reflects contracting performance outcomes such as the availability of […] systems” (Glas et al., 2013)</td>
</tr>
<tr>
<td>3. KPI design and</td>
<td></td>
<td>“Performance measures […] are perhaps the most important elements of performance contracting” (Alyami et al., 2015)</td>
</tr>
<tr>
<td>performance targets</td>
<td></td>
<td>“Once baseline values are available for the selected metrics, the next step is to set objectively the target values for them […]” Sols et al. (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Select KPIs that reflect customer value is often difficult, since the customer’s needs often are formulated in abstract terms” (Holmbom et al., 2014)</td>
</tr>
<tr>
<td>4. Design of incentive</td>
<td></td>
<td>“PBC replaces conventional cost-plus contracts” (Kim et al., 2007)</td>
</tr>
<tr>
<td>system</td>
<td></td>
<td>“[…] the pay-on-availability approach constitutes the basic payment model of PBC in manufacturing industries” (Hypko et al., 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“It is conceivable that all payments fully relate to the performance indicator, but parts of the payment can also form an incentive” (Glas et al., 2013)</td>
</tr>
<tr>
<td>5. Data capture</td>
<td>Contract</td>
<td>“Assessment of outcome quality […] using three data sources: passenger satisfaction surveys, passenger as well as train conductor complaints, and quality audits” (Nullmeier et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>management</td>
<td>“The costs for gathering performance information can outweigh the benefits […]” (Greiling, 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“[…] the evaluation of the influencing factors was not conducted in a systematic and structured way” (Kumar et al., 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“These data are analyzed and subsequently used to provide feedback to the supplier” (Nullmeier et al., 2016)</td>
</tr>
<tr>
<td>6. Performance</td>
<td></td>
<td>“Success is measured against the established metrics that are reported on regularly […]” (Vinue and Manrodt, 2012)</td>
</tr>
<tr>
<td>assessment</td>
<td></td>
<td>“Performance evaluation can be problematic in terms of evaluation routines” (Selviaridis and Wynstra, 2015)</td>
</tr>
</tbody>
</table>
and vice versa, which will be outlined and summarized in this section. The review revealed that the PMM aspects in PBC are practice driven because most PMM aspects are described based on case studies. As mentioned above, references to common PMM knowledge, such as the balanced scorecard (Kaplan and Norton, 1992), are not reported. PMM aspects are addressed and solved practically without actually using the existing PMM knowledge. This aspect is supported by the fact that only four reviewed articles were published in dedicated PMM journals, such as the *International Journal of Productivity and Performance Management*. Therefore, the PBC literature could generally profit from focused studies on the implementation of performance measurement systems in close buyer-supplier relationships (e.g. Van Camp and Braet, 2016; Ferreira et al., 2012). This situation exists in particular with respect to process orientation, because the PMM literature provides several process models for PMM in contractual relationships, e.g., by Neely et al. (1995) and Bourne et al. (2005). The models can be transferred and adapted to PBC, together with other features of a PMM system (e.g. Franco-Santos et al., 2007; Choong, 2013).

Conversely, expanding the PMM literature further to include emerging new practices and business models such as PBC is also required (Melnyk et al., 2014), although there are some examples that analyze PMM with regard to PBC or servitization (e.g. Neely, 2008). Specifically, PBC can contribute to PMM in several ways. First, PBC provides a new contingency setting for PMM, which is highly service and value oriented and aligns buyers with supplier objectives. Second, PBC focuses on outcome-oriented KPIs, which pose several challenges, e.g., how to measure, report and analyze causal relationships between inputs, outputs and outcomes. The PBC literature develops and proposes outcome-oriented KPI systems Sols et al. (2007). Such new and specific findings can be transferred to and cross-validated with existing KPI frameworks from the PMM literature. Third, the PBC literature analyzes incentive mechanisms (Kim et al., 2007) and develops models to explain and structure compensation schemes (Sols et al., 2008; Glas et al., 2013) that can deliver fruitful insights for the PMM literature. Fourth, PBC can contribute to the PMM discussion with insights from new and developing theories, such as the service-dominant logic theory and its PBC implementation (Randall et al., 2010).

To provide more specific guidance on how both literature streams can learn from each other, a brief description of research opportunities in each process step of our analysis framework is provided. This illustrates that a comprehensive design and implementation model for PMM in PBC is missing, as in each step, there are promising opportunities.
Process step 1, alignment with corporate objectives, has not been identified as an important aspect in the design of PBC, although the need for a link from performance measures to corporate strategic objectives is consistently agreed upon in the PMM literature (Keegan et al., 1989; Otley, 1999; Bourne et al., 2005; Franco-Santos et al., 2007). The strategic objectives of an organization must first be considered before the performance expected from a contractor within a PBC scenario is specified (Barber and Parsons, 2009). The examination of corporate and supply relationship objectives, such as quality, innovation, cost, environment, design or even customer satisfaction, is promising, as this is the starting point for PBC design.

In step 2, performance specification, there are still ambiguities, as performance is often described as being output or outcome oriented; however, both terms are not defined in detail. Without knowing what performance really is, it is not possible to optimize the design of the PMM system in a PBC context. The definition and specification of performance in the PBC setting appears to be a promising research task, and it could link to the discourse in the PMM literature that analyzes the similarities in the definition of performance, e.g., multidimensional understanding (Lebas and Euske, 2007) or possible perspectives such as output and outcome but also input and throughput (Entchevmeier, 2008).

Next, the selection of the right KPIs in step 3 is a major topic in the PBC literature, in which KPI systems are developed and advice on how to select or implement KPIs is provided (Sols et al., 2007). This is similar to approaches in the PMM literature (Keegan et al., 1989). Once the relevant KPIs are defined, the required level of performance must be determined (Ferreira and Otley, 2009). Given the link between achieved performance and the compensation of a provider, the setting of performance targets is a critical task within the design of a contract (Ferreira and Otley, 2009). Because many performance-based contracts are individually designed, the ability to develop a standard set of KPIs only seems possible to a limited extent. However, in some branches, such as human services, construction and road maintenance, standard sets of KPIs have been developed.

Step 4, the payment and incentive system, was identified as an important process within the PMM context (Neely et al., 1995; Franco-Santos et al., 2007). The incentive component as a major governance element is a key determinant of PBC and helps to focus the contractor’s attention on the defined performance expectations (Neely et al., 2005). Five payment models have been identified in the reviewed literature that are suitable for PBC. The most urgent challenge in this context is to integrate the right KPIs and to set appropriate performance levels.

Step 5, data capture, is a necessary requirement for calculating KPI values. It appears to be an underestimated task. This is surprising, as in a PBC context, the provider’s performance is at least partially tied to the achievement of performance goals (Selviaridis and Wynstra, 2015), so the gathered data must be precise, accurate and reliable (Choong, 2014). Therefore, real-time data capture that is supported by IT is recommended. Particularly when KPIs and the respective data are part of a contractual relationship, investments in information technology foster reliability, accuracy and trust in the determined KPIs (Forslund, 2012). In the design phase of the contract, the issue of how to measure the selected KPIs must be determined. Only when the data have been gathered as objectively as possible will the contractor be willing to tie a portion of the compensation to the defined KPIs. In summary, the reviewed PBC literature lacks deeper insights into approaches for reliable data gathering.

Next, data must be assessed (step 6). Performance assessment procedures aim to evaluate service success, assign failures to specific actions and investigate who is accountable for failures. The question of how to gather data is almost never discussed in the PBC literature; however, the PMM literature, of course, addresses that step in detail and, e.g., divides
performance assessment into basic and extensive data analysis (Bourne et al., 2005). Basic analysis is performed by an IT system and provides deviations against target, often with a traffic light symbol, and automatically calculated figures are presented as, for example, month-to-date figures. Extensive data analysis is also often provided by the IT system; however, it is time consuming and supported by enhanced analytical IT tools, e.g., data enquiry tools and opportunities for a drill-down of KPIs (Bourne et al., 2005). The PMM literature also suggests initiating regular meetings between contractor and customer to assess and interpret the collected data to foster improvements (Mortensen and Lemoine, 2008). It appears that there is great potential for knowledge transfer from PMM to PBC in that step.

The discussion of process step 7, performance reporting, is still in its infancy. Reporting frequency, report content and design are only rarely mentioned in the PBC literature. In contrast, performance reporting is a powerful instrument for communicating the measured performance, targets and priorities, motivating people, providing a basis for decision making and ensuring that the project is aligned to its objectives (Otley, 1999; Choong, 2014). The PMM literature addresses that issue in detail, which indicates opportunities for future research.

Process step 8, taking action and refreshing, is discussed in the PBC literature only when referring to issues of contract extension and contract flexibility. Other actions are barely addressed, whereas the PMM literature suggests that specific actions might be required in the course of the contract, e.g., actions to ensure that KPIs and performance targets remain relevant, achievable and challenging over the whole contract period (Kennerley and Neely, 2002). Further issues that are discussed in the PMM literature – such as PMM issues that concern who may act and which events require action taking, e.g., dramatically declined performance measures (Bourne et al., 2005) – are not examined. In addition, a periodic re-evaluation of the whole PMM system is recommended in the PMM literature (Bourne et al., 2005; Franco-Santos et al., 2007; Pinheiro de Lima et al., 2013); however, it is not a topic in the examined PBC literature.

After summarizing and discussing the answers of the guiding research question and presenting further insights from the PMM literature for each process step of the analysis framework, aggregated theoretical and managerial implications can be outlined in the next section.

7. Implications

The findings from the literature review and the comparison with insights from other PMM studies in the discussion section indicate several fits and gaps between the two research areas. Both feature performance as a key concept but differ in the profundness of their analysis and discussion. Referring to the analysis framework of this paper, the contract design phase is addressed quite often, and the level of the discussion within the PBC literature is high in some aspects, even when there remains room for further analyses of some issues. Fewer contributions address PMM aspects in the contract management phase or the post-contract phase. The topics are only briefly or simply mentioned. Thus far, the level of academic discussion in these contract phases is relatively low.

The interpretation of the review findings has a number of implications for researchers in the PMM and PBC fields. The implications are presented in the form of four major future research opportunities. These research opportunities show pathways to extend PMM theory in the emerging practical field of PBC.

7.1 Research opportunity 1: strategy alignment

Although the alignment with corporate goals is identified as a major issue in the PMM literature, it is only slightly touched upon in the context of PBC. This finding is astonishing because PBC performance is outcome oriented and thus should be of strategic interest.
It appears that there is conceptual uncertainty with regard to how strategic goals and business outcomes are aligned between the PBC provider and its PBC customer. It is recommended that future research perform explorative and conceptual work on that issue to identify how and why corporate strategic goals, relationship strategic goals and the design of PBC are connected.

7.2 Research opportunity 2: taking action and refreshing
This topic is only addressed marginally, whereas the PMM literature addresses triggers for actions and the causes-effects of actions and improved performance. Therefore, future research can focus on triggers for actions that are required from PBC managers to ensure ongoing performance improvements. Because this topic has hardly been addressed to date, examining the issue exploratively and empirically (e.g. through case studies) is recommended to determine whether there are specific and conceptual differences in the execution of that PMM phase when conducting business in a PBC.

7.3 Research opportunity 3: performance measurement
A detailed discussion of performance monitoring and reporting issues (data capture, performance assessment and performance reporting) is missing in the PBC context; in particular, the approaches for data analysis and cost/benefit issues of automated data gathering should be considered. Due to the numerous publications that concern this topic in the PMM literature, a conceptual study is not suggested; rather, an empirical assessment is suggested that addresses whether a transfer of related topics from the PMM to the PBC field is appropriate. This assessment would include further research on the necessary IT infrastructure and systems that support PBC performance measurement.

7.4 Research opportunity 4: specific aspects of PMM design
The last research opportunity is related to individual aspects of PMM design for PBC because the design of KPIs or incentives is already broadly addressed; however, several challenges remain open for discussion. One of these challenges is the selection of appropriate KPIs that reflect the desired outcome in a PBC. A process description to develop KPIs for a PBC is missing, and further research could examine this specific topic. Another challenge is clarification of the definition of performance. Currently, most articles refer to outcome-oriented performance, while some refer to output-oriented performance. PMM widely acknowledges a multidimensional definition of performance. Therefore, the conceptual refinement of the understanding of performance is suggested. The third specific aspect focuses on the link between PBC KPIs and their integration into the (corporate) PMM indicator system. For that issue, empirical research is recommended.

In addition to the implications for research, this review also proposes managerial implications. More specifically, the process-oriented understanding of PMM aspects helps practitioners to structure their measurement and management activities in accordance with the eight process steps. Furthermore, the content that is presented in each process step is helpful to crosscheck and validate PBC efforts in practice with the state-of-the-art. For example, PBC practitioners are now able to compare their KPIs with the indicators that are listed in the tables of this paper. Thus, the analysis framework can direct management’s attention to crucial process steps, but it also provides initial help for operations management within single process steps. However, some PMM aspects, e.g., the important issue of IT support systems for data capture and reporting, are so new that practice should cooperate with researchers to allow case analysis or best-practice identification studies.
8. Conclusion
This contribution sheds light on the emerging business concept of PBC, which is explicitly concerned with performance and its measurement, because payments are directly linked to KPIs. This review synthesized two literature streams and revealed that, surprisingly, only some aspects of PMM are discussed in the context of PBC, whereas many aspects are only marginally addressed. Of the eight PMM process steps in the analysis framework, only two are discussed in more detail.

The first main contribution lies in the synthesis of PMM and PBC literature and the creation of a combined and cross-understanding that enables researchers and practitioners in both areas to learn from each other. The second main contribution lies in the literature-based identification of future research opportunities. Four opportunities are highlighted, each of which provides inspiration for the PMM and PBC research streams. The findings provide evidence that the PMM and PBC research streams can profit from each other. PMM could further develop in the direction of service-oriented businesses such as PBC. The examination of whether and what knowledge can be transferred to PBC, e.g., in the field of performance reporting, is another recommendation. In a broader sense, this research also contributes to the literature on supply chain integration (e.g. Fawcett and Magnan, 2002; Flynn et al., 2010) because PMM connects the PBC customer with its provider.

This investigation has made efforts to base its findings on robust and rigorous research, i.e., using coding software, involving different researchers in the coding process and assessing inter-rater reliability. However, there are research limitations that must be considered. First, this research work faces all of the limitations of literature reviews in general, such as the need to focus on specific keywords and employ a filter system to reduce the literature hits to a manageable number. Second, content analysis always faces the risk of coding bias and subjectivity. Finally, the interpretation of the findings into future research opportunities involves the aggregating of information.

References


Aspects in performance-based contracting


Further reading


(The Appendix follows overleaf.)
### Table AI
PMM aspects in PBC literature by industry branch

<table>
<thead>
<tr>
<th>Industry</th>
<th>Strategy alignment</th>
<th>Performance specification</th>
<th>KPI design and performance targets</th>
<th>PMM process step</th>
<th>Data capture</th>
<th>Performance assessment</th>
<th>Performance reporting</th>
<th>Taking action and refreshing</th>
<th>Sum/100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerospace/defense</td>
<td>3</td>
<td>5.6%</td>
<td>18</td>
<td>33.3%</td>
<td>12</td>
<td>22.2%</td>
<td>11</td>
<td>20.4%</td>
<td>1 1.9%</td>
</tr>
<tr>
<td>Healthcare/welfare</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>12</td>
<td>32.4%</td>
<td>10</td>
<td>27.0%</td>
<td>7 18.9%</td>
</tr>
<tr>
<td>Construction</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>5</td>
<td>19.2%</td>
<td>8</td>
<td>30.8%</td>
<td>1 3.8%</td>
</tr>
<tr>
<td>Transportation</td>
<td>3</td>
<td>13.6%</td>
<td>4</td>
<td>18.2%</td>
<td>6</td>
<td>27.3%</td>
<td>2</td>
<td>9.1%</td>
<td>1 4.5%</td>
</tr>
<tr>
<td>Energy</td>
<td>1</td>
<td>6.7%</td>
<td>4</td>
<td>26.7%</td>
<td>3</td>
<td>20.0%</td>
<td>0</td>
<td>0.0%</td>
<td>1 6.7%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>1</td>
<td>7.7%</td>
<td>4</td>
<td>30.8%</td>
<td>4</td>
<td>30.8%</td>
<td>3</td>
<td>23.1%</td>
<td>0 0.0%</td>
</tr>
<tr>
<td>Public sector</td>
<td>0</td>
<td>0.0%</td>
<td>4</td>
<td>19.0%</td>
<td>6</td>
<td>28.6%</td>
<td>3</td>
<td>14.3%</td>
<td>4 19.0%</td>
</tr>
<tr>
<td>Housing</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>25.0%</td>
<td>0</td>
<td>0.0%</td>
<td>1 25.0%</td>
</tr>
<tr>
<td>Oil and Gas</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>1 25.0%</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>8.2%</td>
<td>13</td>
<td>21.3%</td>
<td>10</td>
<td>16.4%</td>
<td>6</td>
<td>9.8%</td>
<td>9 14.8%</td>
</tr>
<tr>
<td>Sum</td>
<td>13</td>
<td>5.1%</td>
<td>64</td>
<td>24.9%</td>
<td>60</td>
<td>23.3%</td>
<td>33</td>
<td>12.8%</td>
<td>26 10.1%</td>
</tr>
</tbody>
</table>
About the authors
Andreas H. Glas is an Assistant Professor at the Competence Network Performance Based Logistics at Bundeswehr University Munich. His research investigates the buyer-supplier cooperation and is in particular focusing on performance-based contracts and service-based innovation. He co-edited the book Performance-Based Logistics and is working on several research topics linked to projects in the automotive, manufacturing and aerospace industries. Andreas H. Glas is the corresponding author and can be contacted at: andreas.glas@unibw.de

Florian U. Henne holds an MSc Degree in Business Administration and has been working for an international management consultancy for more than ten years. He has managed numerous projects in the field of procurement, supply chain management and performance management in general. Besides his job as a Management Consultant, he is a Researcher at the Chair for Materials Management and Distribution at the Bundeswehr University Munich. His research is related to performance management aspects of performance-based contracting and their influence on contract success.

Michael Essig holds the Chair for Materials Management and Distribution at the Bundeswehr University Munich and is the Co-Director of the Research Center for Law and Management of Public Procurement as well as the Competence Network Performance Based Logistics. From 2009 to 2012, he acted additionally as the Vice President for Research at Bundeswehr University. His main research interests are strategic supply management, supply networks, supply chain management and public procurement.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Operations flow effectiveness: a systems approach to measuring flow performance

Mohamed Afy-Shararah
Faculty of Environment and Technology, University of the West of England,
Bristol, UK, and
Nicholas Rich
School of Management, Swansea University, Swansea, UK

Abstract
Purpose – Effective operations management systems (OMS) measurement remains a critical issue for theorists and practising managers (Neely, 2005; Bititci et al., 2012). Traditional labor efficiency measures sufficed when all that was made could be sold or when mass production systems filled warehouses with stock and the OMS had little relationship with “the consumer.” Modern manufacturing systems require a different form of flow optimization (beyond labor efficiency) measurement (Schmenner, 2015). The essential unit of measure for all OMS designs is the optimal use of time for process value adding and the flow of materials into and from the conversion process. Timely flow, therefore, satisfies the needs of multiple organizational stakeholders including cash flow (accounting), consumer reaction times (marketing) and the general steady state flow of materials (sales and supply chain). The purpose of this paper is to present the results of testing a new performance measure of operations flow effectiveness (OFE) with ten purposively selected cases.

Design/methodology/approach – The paper is theory building using ten, purposively selected, longitudinal case studies drawn from the UK high-value manufacturing (HVM) sector using a pluralist methodology of interviews, observation and secondary data.

Findings – The OFE measure provides a holistic view of material flow through the input-process-output cycles of a firm. The measure highlights OMS design weaknesses and flow inhibitors that reduce cash flow using a time-based approach to measuring OMS performance. The study validates the OFE measure and has identified six key design elements that enable high flow performance.

Originality/value – The paper tests a new process-focused flow performance measure. The measure supports a holistic approach to the manufacturing enterprise and allows different OMS designs to be evaluated so that organizational learning may be enacted to support performance improvement.

Keywords Performance measurement, Operations strategy, Flow, High-value manufacturing, Case study

Paper type Research paper

Introduction
Modern operations management (OM) discourse, in an era of make to order, smaller batch sizes, and a greater emphasis on waste elimination (Womack and Jones, 1996), has returned to the central question of what is flow and how should it be measured. The traditional dominance of mass production measures – myopically focused on labor productivity and maximum utilization of the production process – has attracted criticism in most volume and repetitive order settings. For the modern business, measuring and managing for pure process efficiency is unlikely to nurture the “externally supportive” position so cherished by Hayes and Wheelwright (1984) nor does such “traditional efficiency” (essentially labor productivity) align well with a modern concern for “servitization,” personalized production, mass customization and the new technological advances presented by Industrie 4.0, 3D Printing and new technology-based “S” curves (Rogers, 2010).

The concept of “system flow” is poorly defined for these new emergent OM models. The increasing emphasis on the circular economy and a migration to higher value adding sectors for mature economies, means productivity-based cost efficiency measurement is increasingly a questionable OM measure (Schulte, 2013; Bonciu, 2014; Lewandowski, 2016). The dysfunctions of applying ineffective or poorly aligned “flow” measures has been well
explored in the literature (Goldratt, 1990; Deming, 2000), however, agreement on effective flow performance measures has yet to be reached for new high-value manufacturing (HVM) contexts and a new world of ultimate flexibility and personalization. The latter may provide manufacturers with access to product markets where greater profits exist from personalized production but it does not allow operations managers to ignore the measurement of process “efficiency” for learning and control (Technology Strategy Board, 2012a; MacBryde et al., 2013). It is not just a “developed” country challenge to transition to higher value manufacturing. Emerging manufacturing economies have already overtaken France and the UK in terms of manufacturing gross value added (GVA) (Mckinsey Global Institute, 2012) so despite the progress made in OM technology and models, the measurement of flow remains a key challenge. The modern imperative is therefore to measure value added time and understand how high performance is supported by organizational features that optimize people, information, equipment, material flows and improvement based on organizational learning to occur (Größler et al., 2006).

This paper focuses on purposively selected cases drawn from the British HVM sector as a context most likely to reflect the general movement from traditional mass production designs to more experimental operations management systems (OMS) models. A cross-comparative case study involving ten manufacturers was selected and the operations flow effectiveness (OFE) measure was applied to assess the flow performance of each organization as a means of validating the measure. Additional assessments of the “states” of six key design elements were undertaken to determine how these “states” support high levels of OM flow performance.

Operations management and measurement
A systematic literature review was undertaken using key word searches and citation databases including EBSCO, Google Scholar, Emerald Insight and Science Direct archives to identify studies of OM performance measures. The keywords used for searches included OM and measurements, supply chain management and measurements, material flow measurement and other synonyms aligned with the subject of flow measurement. The results showed a lack of systematic studies and conceptual studies and a bias of papers toward lean systems reviews, environmental/green and sustainability performance measures, studies of risk measurement or humanitarian logistics performance measures, modeling of specific cases, sector specific measures and benchmarking studies. These studies were reviewed and, whilst useful, most failed to conceptualize material flow. After systematic reduction, only 33 papers, published since year 2001, were considered of use to this study and deemed to meet the qualities of process measures of flow with properties that support flow measurement of firm-level process performance.

Literature review
The review of the extant literature commences with the positioning of OM measures for strategic elevation and the relevance of OM to a modern business. It will highlight the importance of “material flow” and its central relationship with improved cash flow (economic sustainability), and then expose the measurement gap in the modern approach of “swift and even flow” (SEF).

The modern debate concerning OM measures and objective setting (Neely et al., 2005) can be traced to (Skinner, 1969) and the rejection of the cost efficiency “trade-off” where a cost focus compromised all other performance objectives. For Skinner, performance was conceptualized as lowest unit cost and set within the dominant model of scientific mass production. A cost focus led to a focus on productivity through utilization to the point of making products without orders. Later authors, founded their applied models on a “quality first” approach to flow (Nakane, 1986; Ferdows and De Meyer, 1990) which better aligned
with a competitive market environment, differentiation and the corporate sales function. The OMS “sandcone” model (Ferdows and De Meyer, 1990) exemplified this distinct logic and evolutionary approach to performance objective mastery based on quality and treated cost as the outcome of OM design decisions (technology, layout, etc.). The quality focus generated benefits in terms of time compression, dependability and later flexibility of the production process as the competitive weapon for strategic elevation that Skinner (1969) and Hayes and Wheelwright (1984) so desired.

The “quality first” agenda of the 1980s and 1990s onwards, supported the total quality management and later the lean approaches to OMS and extended measurements toward a holistic understanding of end to end performance measurement. The emphasis on quality improvement and waste reduction supported business viability by linking the OMS with better financial performance, greater cash flow and better cash management to avoid insolvency of fundamentally profitable businesses (Boer et al., 2015).

In the quality and lean era of flow measurement, information and material flow were linked to cash flow and the economic success of the firm through a minimal consumption of cash. Unlike the traditional mass production measures that accelerated the consumption of cash through over-production, additional storage requirements and obsolescence costs (Womack and Jones, 1996). The lean approach exposed the apparent traditional low unit production costs as masking a higher overall total cost of production (Ohno, 1988). The lean approach based greatly on Deming (2000) supported the hypothesis that sustainable competitive advantage was founded upon high material flow derived from high quality processes (Womack et al., 1990).

However, lean systems and quality systems rely heavily on repetitive flows of standardized products and a movement from large to small batches then one-piece flow (Womack and Jones, 1996). But these methods are employed in relatively closed and oligopolistic industries where demand can be controlled and products can be standardized. Internal measures of lean systems favor a rhythm (Takt time) to replenishment or measures such as the overall equipment effectiveness of a given asset (Nakajima, 1988) and where production was nonetheless controlled. Indeed, during this era of manufacturing (post mass production) the “Theory of SEF” by Schmenner and Swink (1998) emerged but stopped well short of offering a set of performance measures that can be used to test the effectiveness and viability of lean manufacturing (Gregory, 2007; Rios, 2010). The lean approach did, however, introduce the concept of time compression and just in time management (Monden, 2011) as well as a movement from a current designed OM “steady state” to a future state where OM redesign would enhance flow in a closed system – such as the automotive, aerospace, electronics and food production sectors (Rother and Shook, 2003). The application of lean flow systems in more open systems, where control is less easy to apply proved a challenge and resulted in maintaining excessive capacity to absorb less regular demand or a return to greater inventories of standardized products (Klug, 2013).

SEF therefore rested heavily on enhanced physical flow through a reduction of “noise” in production subsystems (making), and informational exchanges (that trigger production). Where information could not be controlled and departmental measures of utilization were present, local process optimization inhibited the system level of material flow needed for effective cash flow. System “noise” and variations within a lean steady state arise from erratic demand and are amplified by unreliable equipment, erratic supplier performance and inaccurate customer forecasts. They inhibit progress toward stockless make to order production and these variations halt the flow of materials for an entire closely coupled production and supply system. Closed lean systems and SEF seek to reduce the noise of system dynamics (Forrester, 1961; Burbidge, 1985; Towill, 1997; Forza and Salvador, 2001; Childerhouse and Towill, 2003; Geary et al., 2006; Durugbo et al., 2014; Huo et al., 2014) to enhance process flow (Schmenner and Swink, 1998; Schmenner, 2001, 2015; Schmenner et al., 2009). The underlying premise of the
theory is that “the more swift and even the flow of materials through a process, the more productive that process is” (Schmenner and Swink, 1998, p. 102), regardless of the capital intensity or products offered by the firm.

Despite the weaknesses of SEF, it explicitly seeks to maximize “value added” activities whilst minimizing “non-value added” wasteful ones internally (Ohno, 1988). SEF also supports organizational learning and improvement processes using Six Sigma, etc., to reduce such noisy variation (Schroeder et al., 2008). Schmenner and Swink (1998) argue noise and waste elimination result from management decision making and OM working practices and include attempts to dampen it through lean practices (Shingo, 1981), and the poorly researched methodology of total productive maintenance (TPM) (Nakajima, 1988). SEF also seeks bottleneck optimization by reducing process variation and reducing the forms of noise and compromised “laws of variability” identified by Hopp and Spearman (2008) and the earlier works of Goldratt (1990).

At this current juncture in OM thinking, SEF appears appropriate for the design of technical systems for speed and time compression based on the “quality first” approach to optimizing available productive time (minimizing errors/noise) to achieve greater flow performance. However, SEF has significant weaknesses and poorly accounts for modern approaches to manufacturing, the provision of higher value added products and the personalization of goods needed when dealing with modern customers. In short, SEF is effective for relatively closed systems making standardised products. Yet modern conditions call for specialised products made in a much more flexible approach to product delivery, new socio-technical system designs needed to support truly “on demand” manufacturing and a fundamental need to review how flow is measured. This new “post lean” manufacturing domain includes the challenge of Industry 4.0 and cyber-physical systems (Liao et al., 2017). Despite decades of OM thinking, modern OMS advances have exposed fundamental issues with the conceptualization of and “supporting states” of organizational features that support high levels of flow for a firm (Boer et al., 2015). The next section will explore the alignment of performance management systems and the exploitation of flow by an OMS.

Performance management and measurement systems

There is no universally agreed constitution of an effective performance management system (Pinheiro de Lima et al., 2013), yet Franco-Santos et al. (2007) found 17 different definitions for the term. Definitional disagreement has created confusion which limits the generalizability of traditional research (Marr and Schiuma, 2003). Neely et al. (2005) usefully argue that a performance measurement system (PMS) is “the set of metrics used to quantify both the efficiency and effectiveness of actions” (p. 1229), they also argue for a portfolio of measures to be used for modern complex and contextually embedded businesses (Srimai et al., 2011). PMS structure superordinate company goals into measures for control purposes and the measurement of flow. Traditional labor productivity and crude measures of cost efficiency oppose modern priorities for profitable customer satisfaction with minimal delay, maximum personalization at the lowest possible level of waste.

PMS should support and focus OM (Drucker, 1954; Power, 1997), yet it can also become an inhibitor when mismatches exist between corporate need and OM measurement (Neely et al., 2005). Further aligned goals and flow measurement feedback must also be calibrated to meet the speed of the market environment if feedback is to be used to stimulate adaptation of performance and avoid noise. Too frequent feedback creates distortions and triggers reactions in the form of a “permanent sense of crisis” and time-lagged feedback can inhibit flow, particularly cash flow. As such a robust measure of OM resource flow, aligned within a PMS, can optimize flow and learning (Srimai et al., 2011) to once again achieve a swift if not even flow of materials to customers and the achievement of business strategy
In modern times, lean systems may be regarded as achieving quality and delivery of standard products yet the realization of the flexibility stage of the “sandcone” model of mastery and to survive in the harsh modern competitive and pen market requires a fundamental rethink of a PMS and what it promote (Neely et al., 2005).

An effective PMS derived measure will also unite and engage internal stakeholders (Melnyk et al., 2004) thereby reducing the dysfunctions of inconsistency and misalignment of measures between managers (Pinheiro de Lima et al., 2013); authors advocate a top-down and process-focused set of measures to achieve world class performance (Akao, 1991; Jonsson and Lesshammar, 1999; Franco-Santos et al., 2007; Hanson et al., 2011). Alignment of process performance measures should exploit organizational dependencies for collaboration to generate and sustain the performance needed to support the organizational needs. Measures that are closest to the input-process-output cycle therefore support process thinking and support an effective PMS (Table I).

A PMS unites an OMS and the measures of flow performance and poorly aligned systems can sub-optimize business performance. Traditionally PMS favored productivity, profitability, throughput, or quality (Kaynak, 2003; Ahmad and Schroeder, 2003; Koufteros et al., 2005; Merschmann and Thonenmann, 2011), and in the main ignored Neely et al.’s (2005) call for pluralism and “end to end” process approach. Traditional contradictions included achieving optimal labor productivity whilst creating excessive stock, or measures of on time customer service that ignored excessive lead times. Profitability was also a problematic measure that was subject to sales teams (customer price negotiations) and product design process was a measure that could not and typically was not owned by the OM of a firm. Throughput measurements do support flow, yet are often directed only to system bottlenecks management (Goldratt, 1990). Hence, the correct selection of time-based flow measures to support the corporate PMS goals is vitally important to modern businesses.

In a post-lean world, such measures would tend to favor the management of time and responsiveness to customer order fulfillment. The next section will explore the overall equipment effectiveness (OEE), an associated lean production and world class performance measure (Nakajima, 1988) and how it provides a process focused and time-based measure that can also assess the ability of an OM design to handle the flexibility to change between products (McKone et al., 1999) needed for modern open markets. Andersson and Bellgran (2015) argue OEE is one of the most commonly cited OM performance measures in lean operations alongside productivity and has long been used to measure maintenance system performance (Nakajima, 1988) and also production lines (Oechsner et al., 2002; Nachiappan and Anantharaman, 2006). Despite its focus on optimal time-usage for value adding (100 percent OEE), OEE has attracted criticism (Suzuki, 1994; Oechsner et al., 2002; Garza-Reyes et al., 2015). Suzuki (1994) argues operations effectiveness is not purely

<table>
<thead>
<tr>
<th>Features of an effective performance measurement system</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Provides a grounding for communication between stakeholders</td>
<td>Melnyk et al. (2004), Van Aken et al. (2005), Longo and Mura (2008), Cocca and Alberti (2010), Hanson et al. (2011)</td>
</tr>
</tbody>
</table>

Source: The researchers

Table I.
Features of an effective performance measurement system
equipment dependent (materials and labor need to be considered) and Oechsner et al. (2002) propose it should only be applied to individual isolated equipment. Other authors propose inadequacies concerning “green factors,” labor variances and supplier performance (Garza-Reyes et al., 2015). These criticisms are contested by McCarthy and Rich (2015) who claim most criticisms are effectively managed by the measure and its sub-measures see Figure 1. Logistical delays are captured by the asset “availability” and machine speed, availability, process quality, product quality and supply variances are all captured and can be identified by the original OEE measure of Nakajima (1988).

Nakajima (1988) OEE measure evaluate progress toward the “zero loss” and ideal state of TPM. This measure is comprised of three components (Jonsson and Lesshammar, 1999) which are the availability (A), performance (P) and quality (P) – each expressed as a percentage – with OEE expressed by multiplying these three factors together. A resulting OEE therefore indicates performance improvement (Bamber et al., 2003) and it is argued to assist double loop employee learning (Garza-Reyes et al., 2015).

However, research concerning OEE usage (Nachiappan and Anantharaman, 2006; Tsarouhas, 2012; Garza-Reyes et al., 2015) have mainly been case studies and lacked generalization (and were discounted from this study). The authors accepted these criticisms as a failure to holistically measure flow beyond a single asset or tightly coupled production line. It was also perceived to miss the timely provision of inputs and outputs. The contemporary OM requirement is to adopt a “whole process” view that the traditional OEE “point measure” could not provide and any measure would need to be capable of usage in different OM contexts. The OEE measure could be operationalized, with modification, to measure SEF when integrated with input and output quality and delivery performances.

The contemporary context of modern manufacturing is post-lean, it requires rapid and timely order fulfillment of diverse and customized product ranges and takes performance to a higher level of flexibility than previously posited by the “sandcone” model (Ferdows and De Meyer, 1990). SEF has been theoretically attributed to high OM performance (Schmenner and Swink, 1998; Seuring, 2009) but it lacks a measure of holistic flow performance. Given the high priority attached to the context of HVM by developed and developing economies, few studies of flow performance exist, hence this gap became the focus of this theory building study. The HVM sector represents a diverse manufacturing sector with significant variability in products and supply chains that are more representative of the different forms of OM process choice.

Research design
The authors adopt a realist approach in grounding this context-rich study of performance measurement (MacCarthy et al., 2013). The approach allows for theory building and to

---

**Figure 1.** Calculation of overall equipment effectiveness

*Source: The researchers, adapted from Nakajima (1988)*
identify “outliers” whose OM systems defy the norm (Voss et al., 2002; Sousa and Voss, 2008). The HVM sector is under-researched and the poor definition of the subject and this context made the study inappropriate for a positivistic methodology (Stuart et al., 2002). Flynn et al. (1990), Eisenhardt (1991) and Yin (2013) all support the appropriateness of such a methodology as presented in Figure 2.

OM theory has often been derived from actual practice particularly in areas such as quality, process improvement and performance in operations (Safizadeh et al., 1996; Narasimhan and Jayaram, 1998; Gotzamani and Tsiotras, 2001; Radnor and Gosselin, 2005; Größler and Grubner, 2006; Narasimhan et al., 2006; Zotteri and Kalchschmidt, 2007; Bou-Llusar et al., 2009; Khan et al., 2011; Phan et al., 2011; Soni and Kodali, 2012; Jasti and Kodali, 2014). However, much remains to be understood about OM system performance and measures. The case strategy is an accepted approach and a preferred method for complex organizational analysis (Meredith, 1998; Voss et al., 2002; Stuart et al., 2002; Flyvbjerg, 2006; MacCarthy et al., 2013). The case study increases the likelihood of novel theory development (Boer et al., 2015) especially when applied as with a cross-case comparative method (Leonard-Barton, 1990; Voss et al., 2002).

Given the limited number of cases which can usually be studied in a population of HVM businesses, the definition of HVM is drawn from the UK Government’s TSB (now called Innovate UK), and ten purposively selected cases were selected and included a diversity of technologies, products and market sectors as per Flynn et al. (1990) advocate to select cases that appear to violate the proposed theory. The inclusion of cases that disprove theory ironically enriches theory by showing where it is inapplicable (Eisenhardt and Graebner, 2007). Ten cases were selected of which eight were high-value manufacturers and from the two low value manufacturing category (see Figure 4). The cases were drawn from pharmaceutical, low volume vehicles, electrical mechanisms, semiconductor, automotive and industrial products, electrical systems, medical support devices, home furnishings, filtration and ventilation equipment. The purposive selection criteria and justification draws from the recommendations for generalization from a purposive or theoretical sampling (Silverman, 2013) – see Table II.

<table>
<thead>
<tr>
<th>Selection criteria</th>
<th>Main purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-product falls in the TSB criteria of high-value</td>
<td>To test the theory in the context of high-value manufacturing sector</td>
</tr>
<tr>
<td>Manufacturing sites located in the UK</td>
<td>To ensure all cases are operated under British business laws</td>
</tr>
<tr>
<td>End-product falls in the TSB criteria of low value</td>
<td>To understand if the theory holds outside the high-value manufacturing sector as well</td>
</tr>
<tr>
<td>Management structure of + 100 employees</td>
<td>To increase the probability that a formal management structure exists</td>
</tr>
<tr>
<td>Mature site (more than five years)</td>
<td>To assure that a culture and customary practices exist</td>
</tr>
<tr>
<td>No focus on financial performance</td>
<td>To avoid problems concerning the financial performance of the firm</td>
</tr>
</tbody>
</table>

Table II. Purposive case study selection criteria

Source: The researchers

Figure 2.
A strategy for combining laboratory tests with field tests in theory development

Source: Swamidass (1991)
Qualitative data were strengthened by quantitative data (Eisenhardt and Graebner, 2007) drawn from multiple sources and informants (Jick, 1979) and are shown in Table III. The choice of management informants was restricted to management grades as they hold responsibility for designing features that enable or inhibit high OM performance (Table IV).

Research designs and practical considerations create limitations and for this study include the deliberate exclusion of R&D and design functions and the relationship with flow performance. The definition of HVM businesses implies a reasonable spend by the organization on R&D activity (Technology Strategy Board, 2012a), and a theoretical link already exists between high R&D investment and potentially high performance. This link was not tested in this research and nor was the product design and development in general. Moreover, the financial stability of the cases was not undertaken as financing was not seen as a high influence on operational performance (beyond interruptions to material flow, if accounts were unpaid). No such disturbances were found during the research measures of profitability were not used because some businesses were cost centers, some engaged in transfer pricing for international businesses and the vagaries of profit generation.

This study does not compare direct competitors, instead the flow measure is used to test different OMS contexts. The study of direct competitors within the case studies would have

<table>
<thead>
<tr>
<th>Instrument Form</th>
<th>Theory building</th>
<th>Theory testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-structured interviews</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observational methods</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Archival document tools</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Researcher daybook</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Self-completion questionnaire</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Source: The researchers

Table III.
Data collection instruments and research phases

<table>
<thead>
<tr>
<th>Management informant</th>
<th>Reason for selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operations director/Managing director</td>
<td>Responsible for overseeing the whole system and for matching the operations resources with the environmental demands</td>
</tr>
<tr>
<td>Operations/Manufacturing/Production manager</td>
<td>Responsible for the conversion process and ensuring that the production resources are utilized efficiently</td>
</tr>
<tr>
<td>Quality manager</td>
<td>Responsible for the quality of the materials and the end-products</td>
</tr>
<tr>
<td>Maintenance manager</td>
<td>Responsible for the reliability and availability of the conversion process</td>
</tr>
<tr>
<td>Supply chain/Procurement/Purchasing manager</td>
<td>Responsible for ensuring the availability of high quality materials for production on time and in full</td>
</tr>
<tr>
<td>Logistics/Customer service manager</td>
<td>Responsible for delivering the output on time and in full to the customer</td>
</tr>
<tr>
<td>Human resource manager</td>
<td>Responsible for the recruitment, training and development of staff</td>
</tr>
<tr>
<td>Product design and Development/NPI manager</td>
<td>Responsible for the design and development of new products</td>
</tr>
</tbody>
</table>

Source: The researchers

Table IV.
Reasons for selecting the management informants
provided an interesting insight as the environment, technology, supplier base and customers are the same or similar. Thus, the only differentiation would have been the operations design; such a scenario did not exist in this study.

Another limitation, by design, originates from the UK base for all businesses. As such, there is a single national culture involved with this study and affecting the participating cases. Even though the study focuses on the design of operations, a problem facing all operations managers, national culture will have implications in the treatment and extent to which labor can engage in improvement and how far managers will empower the workforce or organize them into teams.

The UK high-value manufacturing sector
The UK HVM sector is critical to national prosperity. UK economic manufacturing output, as measured by GVA, has endured a steady decline for many decades from more than 30 percent in the early 1970s to 10 percent in 2014. Although the relative importance of the sector has reduced it has reached a plateau Figure 3, countries such as Italy, Brazil, South Korea and France, have surpassed the UK’s position despite the UK remaining strong in key industries (including aerospace where it ranks second only to the USA and is the host country to the Head Quarters of two of the top six pharmaceutical companies see (Advanced Institute of Management Research, 2008). It also boasts one of the most productive automotive plants in Europe operated by Nissan in Sunderland. These national assets in sectors including aerospace, automotive, pharmaceuticals and food were identified by the UK Government’s technology strategy board (TSB) – now known as Innovate UK – as critical to the UK’s future viability and strategy for 2025 (Technology Strategy Board, 2012b).

Defining HVM
The Technology Strategy Board (2012a) defined “high-value manufacturing” as “the application of leading-edge technical knowledge and expertise to the creation of products, production processes, and associated services which have strong potential to bring sustainable growth and high economic value to the UK. Activities by high-value manufacturers may stretch from R&D at one end to recycling at the other” (p. 6). However, according to the Advanced Institute of Management Research (2008) high-value manufacturers are defined as “[…] firms that do not compete primarily on cost. Instead they deliver value for one or more of their stakeholder groups by contracting for capability, delivering product/service innovation, establishing process excellence, achieving high brand recognition and/or contributing to a sustainable society” (p. 5). In other words, manufacturing organizations must go beyond the traditional views of production and

![Figure 3. Manufacturing gross value added as percentage of total economy](source: House of Commons Library (2015))
historic models of high performance to exploit value from modern complicated supply networks (Technology Strategy Board, 2008).

MacBryde et al. (2013) argue that HVM is a state that is achieved when firms move away from competing primarily on “cost” and add value through other means, often by services or high efficiency of production processes making high value goods. As such, “there is no simple definition of high-value manufacturing” (Institute for Manufacturing, 2006) albeit the TSB offers a typology of R&D expenditure and potential for economic growth as key dimensions of HVM Figure 4. The sectors include food and drink, marine and other transport, aerospace, automotive, pharmaceuticals, computers, electronics, optical products, chemicals and electrical equipment and these sectors produce high performance products, have significant investment in high technical R&D, and employ highly skilled staff.

Data analysis and findings
To holistically measure materials flow, the OEE measure (collected over a period of 28 weeks) was extended by including the measurement of quality and delivery reliability of system inputs/outputs to enable the measurement of material flow through the whole system Figure 5. The quality and delivery performance of suppliers was based on the average performance of product specific materials for 28 weeks. Customer quality and delivery performance focused on the typical product which utilized the most resources. Quality was measured as a percentage (defects per hundred) and delivery was calibrated to +/-2 hours of when the product was expected for delivery. The authors considered this measure to be more robust and holistic that those adopted by previous studies.

The results of this case performance analysis allowed the classification of low performer, average performer, or high performer to be allocated to each case (Tables V and VI show case performance). Figure 6 shows the material flow performance – the incoming flow, the internal

![Diagram](image-url)
“Input-Process-Output” model:

Overall Equipment Effectiveness:

Operations Flow Effectiveness:

Figure 5. Performance measures used to assess the case studies

Source: The researchers

Table V. Classification of operations systems based on their material flow performance

<table>
<thead>
<tr>
<th>Classification</th>
<th>Low performers</th>
<th>Average performers</th>
<th>High performers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material flow</td>
<td>Less than 50%</td>
<td>Between 50 and 80%</td>
<td>More than 80%</td>
</tr>
</tbody>
</table>

Source: The researchers

Table VI. Performance of the cases for various analysis criteria, ranked by material flow

<table>
<thead>
<tr>
<th>Analysis criteria</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>QD of suppliers</td>
<td>Quality (%)</td>
<td>81</td>
<td>97</td>
<td>80</td>
<td>99</td>
<td>99</td>
<td>96</td>
<td>95</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Delivery (%)</td>
<td>84</td>
<td>85</td>
<td>90</td>
<td>96</td>
<td>99</td>
<td>97</td>
<td>96</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>OEE</td>
<td>Availability (%)</td>
<td>82</td>
<td>83</td>
<td>79</td>
<td>81</td>
<td>85</td>
<td>87</td>
<td>85</td>
<td>92</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Performance (%)</td>
<td>87</td>
<td>84</td>
<td>95</td>
<td>89</td>
<td>88</td>
<td>95</td>
<td>96</td>
<td>88</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Quality (%)</td>
<td>91</td>
<td>91</td>
<td>97</td>
<td>96</td>
<td>97</td>
<td>99</td>
<td>98</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td>QD to customers</td>
<td>Quality (%)</td>
<td>98</td>
<td>99</td>
<td>97</td>
<td>97</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Delivery (%)</td>
<td>82</td>
<td>89</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>90</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

Source: The researchers
flow and the outgoing flow. The data for the high performers show that these organizations perform exceptionally well on all the measures, whereas other organizations do not show a consistent level of performance. As such, the performance of a manufacturing organization is therefore not only dependent on its own operations, but also on the performance of other members in the supply chain. Consequently, the competitive position of each case will depend highly on the “weakest link” in the chain (Seuring, 2009). This suggests that the strategic configuration of the supply chain is a critical task, which comprises supplier selection as well as the distribution among the customers, and the efficient managing of internal operations.

To enable cross-case comparison, data collected were populated into data tables (Miles et al., 2014) to permit a robust cross-case comparison. Axial coding was chosen to reduce the data set and better manage the comparisons to more effectively identify patterns of design features that promoted high as well as “SEF” of material. Using significant wall space to present the cross-case comparison tables it was possible to see “the big picture” and from which, and over eight rounds of data reduction and pattern identification, Figure 7 was generated to show the significant features that support flow.

**Operations strategy**

The importance of a guiding operations strategy, its development and its purpose in relation to the different case studies (shown in Table VII) highlights the significant differences between high and low performing manufacturing organizations. A fundamental finding shows SEF is enabled by adaptiveness and multi-layered feed-forward planning design element, combined with an open systems approach to strategy and with more frequent updates (to cope with and adapt to the competitive and dynamic environment). It also shows the strong involvement and alignment of staff. The uniting bond of the system is a “quality first” approach adopted by the organization toward customers, suppliers and staff and additional competitive priorities such as speed, dependability and flexibility so that the business can deliver on its promises and generate higher flow performance.

**Communications system**

Table VIII shows significant differences between the case studies with regard to their communication system and information sharing (internally and with supply chain partners).
Fundamentally, the long-term plans (accompanied by short cycles of control feedback) provide organizations with clearer predictions of the future changes and enable them to quickly detect changes/react to maintain SEF. Sharing of information by using a variety of methods for different purposes engages everyone in the organization.
and it gives the lowest hierarchical levels a “voice” as well as the opportunity to innovate and contribute to the overall improvement of the organization. In this manner, information exchange and clarity are used to prevent excessive stocks and chaotic production (information has replaced physically stocked products) and it can be argued that that long feed-forward plans, accompanied by short and frequent feedback cycles result in enhanced performance and enables SEF of materials.

*Supply chain relationships*

The performance of supply chain flow is critical to operational performance of quality material flow. Table IX shows partnering with other supply chain members and the sharing of useful information promotes and enhances material flow. Collaborative working improves flow by sharing the benefits of these improvements of improvements and across the whole chain. Selecting the suppliers and customers with the right capabilities is clearly shown as vital to partnership and uninterrupted material flow. Finally, the maintenance of an alternative and capable supply base ensures risks are reduced and flow can be maintained in the event of a catastrophe at a main supplier.

*Technical system*

The best technical system will never be optimized without robust controls to ensure line availability and quality of the machine, process and material flow. Table X shows a series of control mechanisms including staff engagement and double loop learning are employed at and beyond the focal case organization. These simple devices and artefacts allow operators to escalate concerns and prompt greater responsiveness to restore flow production (they heighten the situational awareness of all staff to flow underperformance and abnormal variance) and thereby prevent waste and interruption. Proactive asset maintenance guarantees availability and production speeds which ensures production plans can be achieved and reduces management noise and re-planning. An effective technical system design is therefore found to be critical to swift and even material flow.

*Social system*

A culture of teamwork and heightened interpersonal dependency is a foundation for modern dominant OM models – but there are significant variances in the range and focus of the skill-sets between workers in organizations of different performance classifications. The high performing organizations rely on staff trained in determining variation in delivery and speed as well as quality variation detection (this allows better use of management as noise is reduced). The low performers use only quality variation detection methods and are unaware of asset underperformance from (operating at lower speeds). The finding shows the critical role that workers have on controlling the environment, the machinery and the quality and a pivotal aspect of SEF of materials. Investments and improving the social system is a major investment by high flow performers acting as learning organizations (Table XI).

*Learning and improvement system*

Staff Knowledge of the technical system promotes learning and improvement of efficiency and effectiveness of OM and Table XII shows the relationship between learning and flow improvement. High performing organizations reveal high organizational learning capabilities reflecting that adaptation is a desired outcome of learning especially in ever changing fiercely competitive markets. Being “fittest” not only involves continuous improvement but a greater integration of worker and machine to achieve greater levels of technical competence and mastery – which again heightens situational awareness of staff.
Table IX. Main findings between high and low performance in the supply chain relationships

<table>
<thead>
<tr>
<th>Performance of organization</th>
<th>Process improvements</th>
<th>Benefits and blame</th>
<th>Sharing of information</th>
<th>Relationship duration</th>
<th>Relationship strength</th>
<th>Selection of supply chain members</th>
<th>Alternative supplier base</th>
<th>Deliveries</th>
<th>Customer service</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Present</td>
<td>Shared</td>
<td>High</td>
<td>Long</td>
<td>Partners</td>
<td>Present</td>
<td>Present</td>
<td>Frequent</td>
<td>Protected</td>
</tr>
<tr>
<td>Low</td>
<td>Limited</td>
<td>Private</td>
<td>Low</td>
<td>Long</td>
<td>Close</td>
<td>Limited</td>
<td>Absent</td>
<td>Infrequent</td>
<td>Loose</td>
</tr>
</tbody>
</table>

Source: The researchers
### Table X. Main findings between high and low performance in the technical system

<table>
<thead>
<tr>
<th>Performance of organization</th>
<th>Production mode</th>
<th>Maintenance units</th>
<th>Systematic breakdown management</th>
<th>Workplace organization</th>
<th>Process standardization</th>
<th>Quality assurance</th>
<th>Design of equipment</th>
<th>Concurrent engineering</th>
<th>Design for manufacturability</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Pull</td>
<td>Proactive</td>
<td>Present</td>
<td>High</td>
<td>All</td>
<td>Total</td>
<td>Present</td>
<td>Present</td>
<td>Present</td>
</tr>
<tr>
<td>Low</td>
<td>Push</td>
<td>Reactive</td>
<td>Absent</td>
<td>Basic</td>
<td>Most</td>
<td>Limited</td>
<td>Absent</td>
<td>Limited</td>
<td>Limited</td>
</tr>
</tbody>
</table>

**Source:** The researchers
<table>
<thead>
<tr>
<th>Performance of organization</th>
<th>Teamwork</th>
<th>Skills set</th>
<th>Product quality skills</th>
<th>Diagnostic skills</th>
<th>Ownership of production process</th>
<th>Machinery skills</th>
<th>Environmental skills</th>
<th>Innovation</th>
<th>Learning capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Present</td>
<td>Multiple</td>
<td>Same</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Enabled</td>
<td>Improved</td>
</tr>
<tr>
<td>Low</td>
<td>Present</td>
<td>Limited</td>
<td>Same</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Restrained</td>
<td>Basic</td>
</tr>
</tbody>
</table>

*Source:* The researchers
to abnormal processing conditions and so to prompt a quicker countermeasure to
maintain SEF.

The key to OM system viability and cash flow management, the primary reason for the Just In Time system (renamed “lean” by Womack et al. (1990) and an underpinning feature of agile systems, is SEF of materials to customers and a creation of a dependent system whereby flows are uninterrupted and aligned. As production batches head toward personalized production in search of higher value adding then it is important to conceptualize flow in a means that reinforces a process view of the firm. The findings show an interesting insight into the design of high and low flow performance businesses operating in modern dynamic conditions.

Discussion

The concept of swift and even material flow is intriguing but lacks a measure of flow performance that is holistic and can be used to learn which aspects of OM design and support structures improve the value adding of production time, and more importantly, flow between the organization and customers/suppliers. Most organizations have ignored Neely’s (2005) call for pluralist measures and legacy measures from the mass production era (cost not value focused) remain to generate dysfunctional behavior and stresses. The OFE measure was found to provide meaningful measurement of supplier performance (on time and in quality) with internal flows through processing lines and the onward quality and delivery performance to customers – a process that mimics cash flow and viability of the OMS. Stocks within and between organizations exist to balance flow or compensate for variations in forecasts, etc. (reflecting a risk appetite for a business). The new flow measure based on measuring at a higher level (“Flexibility level”) of the “sandcone” model (Ferdows and De Meyer, 1990) is uncompromising as the perfect 100 percent score is truly optimal for a business and its existing social and technical features. The measure also identifies the weakest design elements (supply, processing or distribution) and features that need to be improved to move closer to optimal levels of flow (greater availability, cycle time speeds, process quality, etc.) so that learning, mastery and commercial gain results. Moreover, the OFE measures the typical disturbance effects commonly found in manufacturing organizations and combats the weaknesses of the SEF approach (Table XIII).

The lowest performing flow cases reveal a poor cohesion between operations strategy and the effectiveness of other OM features (social, technical and supply chain relationships) and reinforces the need to design rather than emulate. Poor integration, PMS alignment and poor learning result in poor cash flow management (the modal reason for exhausting cash flow and business viability) resulting in a competitive disadvantage. The study finds no individual design element could be identified out as the main contributor and sole responsibility for the overall flow ranking performance and this again reflects the need for a holistic and contingent OM design for flow. High performance businesses operate high flow and highly integrated systems where dependency is high between all direct and indirect functions. The latter “supportive” internal and external environment of customer/supplier relations supports Hayes and Wheelwright’s (1984) proposition that OMS is as competitive weapon even when tightly coupled.

<table>
<thead>
<tr>
<th>Performance of organization</th>
<th>Improvement culture</th>
<th>Problem solving and learning</th>
<th>Improvement across the supply chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Strong</td>
<td>Continuous</td>
<td>High</td>
</tr>
<tr>
<td>Low</td>
<td>Weak</td>
<td>Irregular</td>
<td>Low</td>
</tr>
</tbody>
</table>

Source: The researchers

Table XII.
Main findings between high and low performance in the learning and improvement system.
Disturbances to the production system, to the input and/or output linkages, come in the form of machine breakdowns, tool wear, absent workers, poor information systems and such like. Such inconsistencies and variation in speeds and delivery schedules affects the of utilization of machines or availability of materials. In general production system disturbances to and from the case study are due to delays or incorrect orders and deliveries which confirm the views of (Golinska et al., 2011). Within a case, poor quality is the main cause of disturbance to material flow and such forms of “operational noise” all reduce the physical and information triggered material flow. Such “operational noise” can therefore be defined as “any event, process, or activity that creates excess errors, delays, and/or rework as a result of uncertainties caused by poor information exchange or physical processing.” Operational noise is a main cause for poor SEF and poor material flow which is amplified by:

- ineffective design/redesign decisions that do not support the business in achieving its goals;
- problems and loss of productive capacity;
- problems associated with inbound supply; and
- lack of or inconsistencies in quality and quantity of information exchange.

A variety of strategies were used to manage the dysfunctional impacts of operational noise including dampening techniques such as system buffering (investments in stock to ensure flow but at a cost). Buffering techniques represent a resource-based approach as they require excess materials, machinery, or labor capacity and dampening techniques include information-based approaches based on planning methodologies such as smoothing or physical buffering (Golinska et al., 2011). The study finds an absorptive capacity was created to cope with noise at the high performing cases and a design that, with heightened dependencies between businesses and within the production process improves resilience (Simangunsong et al., 2012). The high performing cases maintained an absorptive capacity rather than attempting “agile” operating model and had proactively managed risks to flow (Scott, 2003). The flow measure was sensitive enough to detect issues with tightly coupled businesses and the need for stock to maintain resilience.

The study finds the most correlated features to high performance (and swift even flow of materials) are a holistic and adaptive multi-layered feed-forward planning system. The study finds higher performers had a more robust, frequent and sophisticated number and portfolio of future planning activities which were managed to ensure the alignment of all the other systems/organizations for flow. Each of the highest performing cases was synchronized through rich and effective communications in short cycles of control feedback accompanied by long feed-forward plans (allowing process managers and trading

<table>
<thead>
<tr>
<th>Typical disturbances effects</th>
<th>Examples</th>
<th>Performance measure</th>
<th>Performance objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material shortages</td>
<td>Delays in deliveries, Incorrect materials or components</td>
<td>Delivery dependability</td>
<td>Dependability</td>
</tr>
<tr>
<td>Breakdowns</td>
<td>Unplanned maintenance, Employee absenteeism</td>
<td>Availability</td>
<td>Flexibility</td>
</tr>
<tr>
<td>Reworks</td>
<td>Poor product quality</td>
<td>Quality</td>
<td>Quality</td>
</tr>
<tr>
<td>Non-standard production</td>
<td>Variability in manufacturing lead time, Variability in delivery times</td>
<td>Speed</td>
<td>Speed</td>
</tr>
</tbody>
</table>

**Table XIII.**

Typical disturbances in production systems, their performance measure, and performance objectives

*Source: The researchers*
partners to look forward whilst controlling short run process variations). The attention to communication system design for the right information at the right time (vertically in both directions and horizontally across the supply chain) is a major distinguishing feature of a higher performing case and interestingly (even allowing for credit terms) for cash flow. It is surmised that higher performing cases are less threatened by bankruptcy due to poor cash flow. Partner relationship management and longevity of trading relationship were equally found at high performing cases which suggest a “fit” between companies and the use of collaboration to remove power struggles and commercial gaming. It is the role of the technical system to physically “convert” the materials that arrive from the supplier into products for the customer. The technical system is designed to have processes that can ensure a SEF of materials. As such, it requires a long-term plan and strict control mechanisms to ensure that the materials do not vary in quality and that the machines are maintained to a good condition and available when needed. The control of the technical system is the responsibility of the workforce in the social system. Therefore, the workforce is empowered and trained to detect any variation in quality, speed and availability that will affect the flow of materials, and to take appropriate action to rectify the problem. The social system is people-focused and is multi-skilled to perform various tasks (and absorb product variety to support SEF and avoid the rigidities of staff with limited flexibility and poor learning/adaptation skills). Investment in the human resources of a business in this manner also increases the ability of staff to detect abnormal variation and respond more quickly to such variation to restore flow. Based on the case findings, Figure 8 is presented to show the holistic properties of a higher performing business and how there is integration of six major design elements that are needed for high performance operations and swift and even material flow.

The authors consider the new flow measure a significant and valuable addition to understanding SEF – including the understanding of swift flow in dynamic and uneven market environments. The ability to establish a 100 percent optimal score for flow (given investments in stock) allows the higher level of the “sandcone” model to be assessed and enacted. The research also identifies the six major OM design elements that enable better flow and support higher levels of operations mastery. The measure therefore offers significant insight into the states and features of OM design that allow such mastery to be effectively exploited in more turbulent product markets as high-value manufacturers take the next steps to change technology, increase product variety and reduce product lives to a single bespoke piece. The new flow measure therefore has utility for most manufacturers and a potential new insight into the OM designs that support Industrie 4.0.

**Conclusions**

The purpose of this research paper was to develop and test a flow performance measurement and to extend the concept of “SEF,” under current dynamic conditions, using the UK HVM sector to theory build from ten longitudinal case studies. The research validates the measurement approach and identifies six design elements that are necessary for the achievement of high performance. The new manufacturing era demands greater flexibility of non-standardized products and has evolved from the standardized and limited product ranges associated with more traditional lean systems. The HVM sector was found to operate at high levels of flow using the new measure and that this form of time management could be applied to all cases. The new era of manufacturing places greater emphasis on the management of time as a proxy measure for cash flow. Traditional measures will come under greater criticism as new models, such as Industrie 4.0, place greater stresses on manufacturers for a quality and tailored product fulfillment process. There remain many issues to overcome with modern manufacturing models including skills needed, ability of the supply chain to support personalized production and the redistribution
of the manufacturing facility itself. The poor state of current knowledge concerning the measurement of flow for “cutting edge” business models of “make on demand” or “make to order” operations requires a new measure of flow to align operations with the business PMS. A reversion to old and “tried and tested” measures, such as labor productivity, will prove both frustrating and will create dysfunctional behaviors. Under new operations models, labor is likely to have comparatively higher technical skills and the OMS will reflect capacity and time utilization when it is needed.

The ten case study manufacturers, ranging in flow performance, were investigated to reveal the OMS design features that supported or inhibited high material flow. The flow measure was developed by the researchers to close a gap in the literature where, despite the acknowledged importance of effective and aligned performance measures (Neely et al., 2005), no real holistic measures exist to support the modern operating models and OMS. This study reveals six design elements that are necessary to achieve high performance including

![Diagram of operational policy deployment and open collaborative partnership.](image-url)
a holistically aligned operations strategy, a synchronizing communications system, a partner-like supply chain relationship, a controlled technical system, an empowered social system and an adaptive and evolving learning and improvement system. These “states” support effective flow management.

The findings emphasize a contingent design of OM features, as no individual design element was identified as the source of higher performance. The study reveals new research gaps in terms of which contingent features (and their sequence of implementation) support new models of OM. The modern manufacturing era will place greater emphasis on learning and the adaptation of socio-technical OM features to enhance flow reliability and responsiveness. A holistic approach to OMS design is required if the enablers of swift flow are to be exploited. Emulation of high performance OMS designs, which has long dominated most lean approaches to manufacturing is no longer a viable option. Such contingent OMS designs are important in dynamic markets where customer satisfaction relies upon flow systems that are supported by a strategic focus of resources, IT enabled decision making and a learning capability to ensure a broad range of personalized products are delivered to customers. Modern organizations must operate multi-layered levels of control at the worker, at the process, and at the supply chain to detect abnormalities quicker and react to control flow. The new measurement of flow, investigated by this study, strengthens the role of OMS design in designing, implementing and supporting high performance operations as a contemporary “competitive weapon.”

The most important activity for operations managers is to maintain the viability of a business and to manage required production time for value added and revenue earning purposes. Figure 8 is a representation of the essential design elements that need to be configured in order to exploit the capabilities of an OMS design for high flow performance. In the absence of “role models” played by such companies as Toyota in the lean era, operations managers must align working practices with the highest levels of flow and do this as a true “end to end” measure. The testing of the OFE measure was conducted with high-value manufacturers that are under pressure to customize and minimize batch sizes, and to improve the generalizability of the measure and approach, other manufacturing contexts will be targeted where product variety is even greater and product life may be as limited as a single piece. Industrie 4.0 (Internet of Things, 3D Printing, Augmented Reality, etc.) provides an ideal opportunity to test the flow measure for swift and uneven flow. It is hoped that this paper will stimulate debate as the cases of this research are continued to form longitudinal studies of change and ideally joined by international partners to test cultural influences on OMS designs and flow performance.

References


About the authors
Mohamed Afy-Shararah is Industrial Engineer and Knowledge Transfer Associate helping businesses implement operational excellence and digital manufacturing to increase productivity and streamline material flow in operations using a variety of approaches from Lean Manufacturing to Systems Thinking. Currently, he is leading a business transformational project to deliver a step-change in productivity through embedding new capabilities in innovative manufacturing techniques and cutting-edge engineering product design. He holds a PhD in Strategic Operations Management from the UK, a Masters in Lean Manufacturing from Egypt and he has also studied System Dynamics in Norway. Mohamed has provided services in lean manufacturing to a diverse range of manufacturing businesses globally. He was Assistant Lecturer at the Industrial and Management Engineering Department in his hometown Alexandria, Egypt. Mohamed Afy-Shararah is the corresponding author and can be contacted at: mohamed.shararah@uwe.ac.uk

Nicholas Rich is Professor of Socio-Technical Systems at Swansea University. He started his career in Cardiff University with Professor Dan Jones (author of the best-selling publications *The Machine That Changed the World and Lean Thinking*). He was awarded the Toyota Fellowship and was trained in Japan. He was co-founder of the Lean Enterprise Research Centre at Cardiff and whilst there has been awarded significant research funding (including patronage by the EPSRC). He has written some of the most downloaded papers in the area of operations management and has written six books. He was Chief Engineer of the 2012 Olympic and Paralympic Medal programme. He joined Swansea in 2013 to continue his research studies of high performance and highly reliable organizations.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
The performance effects of complementary management control mechanisms

Henrik Nielsen and Thomas Borup Kristensen
Department of Business and Management, Aalborg University, Aalborg, Denmark, and
Lawrence P. Grasso
Department of Accounting, School of Business, Central Connecticut State University, New Britain, Connecticut, USA

Abstract
Purpose – The purpose of this paper is to study management control mechanisms (social, behavioral, and output control mechanisms) and their complementary effects on firm performance in lean manufacturing firms.
Design/methodology/approach – The study uses second-order structural equation modeling to analyze survey data from 368 different lean manufacturing facilities.
Findings – The paper finds that the complementary effects of management control mechanisms in lean manufacturing firms outweigh their additive effects on firm performance.
Research limitations/implications – Applying isolated lean management control mechanisms leads to inferior performance, as these management control mechanisms are complementary. Thus, to realize the full potential of lean manufacturing, this paper suggests that lean management control mechanisms should be implemented as an integrated control system.
Practical implications – Firms seeking to benefit from the implementation of lean manufacturing should understand the complementarity among the management control mechanisms, as the performance effects of lean management control mechanisms when applied together are greater than their isolated additive effects.
Originality/value – This paper is the first to provide empirical evidence of the superior firm performance effects of complementary lean management control mechanisms compared with their additive effects. This paper also expands the understanding of how to conceptualize lean management control mechanisms. Specifically, this is the first paper to distinguish between social cultural control and social visual control mechanisms as well as between non-financial and financial control mechanisms. This paper is also the first to use a second-order structural equation model to properly test and account for the complementary effects on firm performance that stem from multiple control mechanisms.
Keywords Lean manufacturing, Firm performance, Complementarity, Second-order model, Management control mechanisms
Paper type Research paper

1. Introduction
Interest in complementarity and in its role in the design of organizations has garnered increasing attention in the academic literature (Ennen and Richter, 2010). Practices that work together are considered to be complementary when doing more of one practice increases the marginal return of another practice and vice versa (Milgrom and Roberts, 1995). Lean manufacturing is an ideal setting in which to study complementarity (Furlan et al., 2011) as it is recognized as an enterprise-wide management system consisting of interdependent practices (Roberts, 2004; Shah and Ward, 2007). Lean manufacturing was conceptualized by Krafcik et al. (1988), when studying Toyota as part of the MIT International Motor Vehicle Program, and it is generally accepted that lean manufacturing improves firm performance (e.g. Fullerton and Wempe, 2009; Hofer et al., 2012; Jayaram et al., 2010; Maiga and Jacobs, 2008).

The authors thank Rosemary Fullerton and Thomas Tyson for their input on survey development and data collection. The three anonymous reviewers and the editor are also warmly thanked for their valuable comments.
However, both Shah and Ward (2003) and Furlan et al. (2010) suggested that it is the simultaneous, systematic implementation of several practices that contributes to firm performance through the complementary effects of these practices. This implies that the partial implementation of practices or of practices that do not work in concert will contribute to a lesser extent to firm performance.

The implementation of lean manufacturing has been found to be associated with companies’ management control mechanisms[1] (e.g. Åhlström and Karlsson, 1996; Fullerton et al., 2013; Kristensen and Israelsen, 2014; Netland et al., 2015), and it is recognized that management control mechanisms can either hinder or help lean manufacturing implementations (Åhlström and Karlsson, 1996; Fullerton et al., 2014). However, there is still much to understand about how management control mechanisms work in the lean manufacturing context. In this study, we investigate the complementary effects of management control mechanisms[2] on firm performance in lean manufacturing companies.

As it is imperative that we examine these management control mechanisms from a holistic perspective (Ennen and Richter, 2010), we utilize the conceptual framework developed by Kennedy and Widener (2008), who extended the work of Ouchi (1978, 1979) and Snell (1992) to management control mechanisms in lean manufacturing companies. Kennedy and Widener’s (2008) framework views management control as interdependent mechanisms consisting of training, visualization, empowerment, peer pressure (social control mechanisms), standardization of practices and rules (behavioral control mechanisms), and performance measurements (output control mechanisms). We extend social management control mechanisms to also include lean thinking (Emiliani et al., 2003), as it is an important catalyst for successful lean manufacturing implementation, and we increase the granularity of Kennedy and Widener’s (2008) framework by distinguishing between social cultural control and social visual control mechanisms as well as between non-financial and financial control mechanisms.

Different strategies are used when testing for complementarity between organizational variables. Ennen and Richter (2010) described two strategies: the interaction strategy, focusing on the complementarity of two organizational variables, and the systems strategy, focusing on the complementarity of a broader set of variables. Using a sample of 368 American lean manufacturing facilities, we adapt the systems strategy and follow the procedure developed by Tanriverdi and Venkatraman (2005). We develop and compare two competing structural equation models: the first model utilizes a second-order factor to capture multilateral interactions and covariance among the management control mechanisms as well as the effects of the second-order factor on firm performance. The second model conceptualizes the management control mechanisms as first-order factors and explores their additive effects on performance.

This study makes two major contributions to the small body of knowledge on this topic. First, we find that the performance effects of a complementary set of management control mechanisms are superior to their isolated additive effects. In fact, three of five management control mechanisms – visual social control mechanisms, financial output control mechanisms, and non-financial output control mechanisms – do not additively contribute to firm performance. Second, our study is the first to provide empirical support from a large sample of firms suggesting that the full set of lean management control mechanisms is complementary. Moreover, we provide detailed descriptions of how lean management control mechanisms work together in order to facilitate a deeper understanding of the complementarity effects on firm performance. We are especially motivated by Fullerton et al. (2013), who call for an extension of their study to encompass all the management control mechanisms from the Kennedy and Widener’s (2008) framework, and by Malmi and Brown (2008), who welcome research on more specified management control mechanisms.
The remainder of this paper is organized as follows: in Section 2, we describe the literature and develop our two competing hypotheses. In Section 3, we present our sample and methods and, in Section 4, we present our results. We discuss and conclude the paper in Section 5. Limitations and recommendations for future research are presented in Section 6.

2. Literature and hypotheses development

It is well established that lean manufacturing is positively associated with firm performance (e.g. Hofer et al., 2012; Fullerton et al., 2014; Khanchanapong et al., 2014; Maiga and Jacobs, 2008). Hence, the focus here is not on whether lean manufacturing can benefit performance but rather on how management control mechanisms assist lean manufacturing companies in achieving improved firm performance. Management control mechanisms have garnered attention in the lean manufacturing literature (e.g. Fullerton et al., 2013; Netland et al., 2015) and have been conceptualized as consisting of social, behavioral, and output controls (Kennedy and Widener, 2008). Research has found that lean manufacturing is related to these management control mechanisms. For example, lean manufacturing has been found to be related to visualization (Banker et al., 1993), peer pressure (Ezzamel and Willmott, 1998), employee empowerment (Lind, 2001), and training (Woolson and Husar, 1998). Lean manufacturing has also been found to be related to standard operating procedures (Rondeau et al., 2000) and rules (Shah and Ward, 2003). Evidence also suggests that lean manufacturing relies on non-financial performance measurements (Banker et al., 1993) and financial performance measurements (Emiliani et al., 2003). Table I depicts the management control mechanisms used in this study. These are drawn from Kennedy and Widener’s (2008) framework, but we increase the granularity of the framework as we distinguish non-financial control mechanisms from financial control mechanisms as well as social cultural control mechanisms from social visual control mechanisms.

Empirical research suggests that lean management control mechanisms are interrelated, but there is limited evidence of their complementarity. For example, in their case study of a lean manufacturing company, Kennedy and Widener (2008) found that social, behavioral, and output controls were interrelated, meaning that, for example, performance measurements (output control mechanism) went hand in hand with employee empowerment (a social control mechanism), and standard operating procedures (behavioral control mechanism), similarly, went hand in hand with visualization (social control mechanism). Kristensen and Israelsen (2014) studied balance among social control mechanisms, behavioral control mechanisms, and output control mechanisms in a single firm. Their results indicated that greater balance led to greater firm performance, and they argued that the results were evidence of complementarity. However, their methodology made it difficult to capture patterns of interactions and covariance among the lean control mechanisms because the control mechanisms were collapsed into two aggregate measures. Without using the management control mechanism terminology, Emiliani et al. (2003) found that social, behavioral, and output controls were interrelated in a lean manufacturing company. Emiliani et al. (2003),

<table>
<thead>
<tr>
<th>Social controls</th>
<th>Behavioral controls</th>
<th>Output controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social cultural controls</td>
<td>Social visual controls</td>
<td>Standard operating procedures</td>
</tr>
<tr>
<td>Employee empowerment</td>
<td>Visualization</td>
<td>Non-financial output controls</td>
</tr>
<tr>
<td>Peer pressure</td>
<td>Rules</td>
<td>Financial output controls</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td>Non-financial performance measurements</td>
</tr>
<tr>
<td>Lean thinking</td>
<td></td>
<td>Financial performance measurements</td>
</tr>
</tbody>
</table>

Table I. Lean management control mechanisms
Kennedy and Widener (2008), and Kristensen and Israelsen (2014) were single firm studies, which makes their findings difficult to generalize. Furthermore, Emiliani et al. (2003) and Kennedy and Widener (2008) did not study the complementary effects of the management control mechanisms on firm performance. In a cross-sectional study, Fullerton et al. (2013) investigated fragmented parts of the lean management control mechanisms. They found that employee empowerment (social control mechanism) and visual performance information (output control mechanism) were interrelated. Fullerton et al. (2013) did not study the complementary effects on performance, and their reductionist method is problematic when studying complementarity (Ennen and Richter, 2010).

To establish clear evidence of complementarity among lean management control mechanisms, firm performance effects stemming from individual management control mechanisms must be compared with performance effects stemming from complementarity of the complete set of management control mechanisms (Tanriverdi and Venkatraman, 2005). Furthermore, a detailed exploration of how the interrelatedness and complementarity of management control mechanisms can support lean manufacturing companies (Maskell et al., 2012) is needed in a cross-sectional setting (Kennedy and Widener, 2008). As we will explain in the sections below, we expect that lean management control mechanisms are complementary and that the complementary effects on firm performance are greater than the additive effects from management control mechanisms. We follow the same argumentation logic and structure as Tanriverdi and Venkatraman (2005). First, in Sections 2.1-2.3, we describe lean management control mechanisms and explain how management control mechanisms are interrelated; second, in Section 2.4, we develop our hypotheses and describe how we expect complementarity to exist between management control mechanisms.

2.1 Social control mechanisms
According to Kennedy and Widener (2008), social control mechanisms in lean manufacturing companies encompass visualization, peer pressure, training, and employee empowerment. Visualization is essential in lean manufacturing companies (Belekoukias et al., 2014; Cunningham and Fiume, 2003), and it goes hand in hand with both behavioral and output control mechanisms. Boards are used in the manufacturing area to visualize the current and future state of operations (a non-financial output control mechanism) and to show standard operating procedures (a behavioral control mechanism). Boards also show whether current activities are deviating from standards (Emiliani et al., 2003) and provide real-time, easy-to-understand performance metrics that direct employees’ attention to potential improvement areas and manufacturing-related problems, ensuring that production objectives are aligned with the lean strategy (Liker, 2004). Training matrices and employee capabilities indicators are used to highlight the skills required for working in a manufacturing cell and to show the current skills for each individual employee working in that cell (Kennedy and Widener, 2008; Maskell et al., 2012). This assists employees during the planning of their work activities. However, visualization goes beyond informing employees about standards, improvement potential, performance, and skills: visualization also includes a structuring of the entire manufacturing area with high visibility, which should allow employees to assist one another between work processes and to help them understand how their own work activities are related to other areas of the facility (Liker, 2004). This can be referred to as global transparency (Adler and Borys, 1996). Global transparency reduces the risk of sub-optimization and enables employees to identify problems and improvement potentials in other manufacturing cells than their own.

For visualization to be effective, employees in lean manufacturing companies must be trained in lean principles (Fullerton et al., 2013) such as kaizen, standard operating procedures, and creativity. Employees not trained in lean principles will not be able to fully grasp, act, and
react to the information on the boards or to use this information to solve problems and identify potential improvement areas. The lean training can be done onsite, e.g. by employees continuously going to the gemba and figuring out solutions or improvements (Farris et al., 2009). Employees are motivated to undergo training, as cell capability indicators highlight whether they are experts in a certain skill (Kennedy and Widener, 2008). The training also facilitates the empowerment of employees responsible for quality, cost, and flow, enabling them to make timely and effective decisions and adjustments to their work (Cua et al., 2001; Fullerton et al., 2013). This is especially important in lean manufacturing companies with reduced buffer inventories, as potential breakdowns have severe effects downstream (Callen et al., 2005; Kristensen and Israelsen, 2014). Additionally, the empowerment of employees enables them to carry out experiments and perform continuous improvement, potentially improving their own and others’ work processes. This, of course, is not something that happens without employees being motivated or being encouraged to do so. A possible motivational element is that lean thinking permeates the minds of employees and managers. Lean thinking enables them to think, act, and behave with a passion for lean manufacturing (Wood et al., 2015), and it therefore functions as an internal motivational factor (Bhamu and Sangwan, 2014). Here, we extend Kennedy and Widener’s (2008) framework, inspired by clan controls (Ouchi, 1979). Peer pressure is another catalyst for employees to solve problems, identify improvement potentials, and undergo additional training (Kennedy and Widener, 2008). Peer pressure in lean manufacturing companies can occur when employees at the same hierarchical level mutually reinforce their desire to obtain additional knowledge, work skills, and higher performance, both in comparison to other employees in the manufacturing cell as well as in comparison to other manufacturing cells and value streams. The monitoring and highlighting of skills and performance within and between manufacturing cells (a non-financial output control mechanism) can lead to a sense of pride among employees and can improve motivation (Kennedy and Widener, 2008). We have decided to distinguish between social cultural control mechanisms and social visual control mechanisms, because the former is input oriented, intended to affect behavior ex ante, whereas the latter is process oriented, intended to guide immediate behavior.

2.2 Behavioral control mechanisms
Behavioral control mechanisms in lean manufacturing companies consist of standard operating procedures and rules (Kennedy and Widener, 2008). These are seen as an aid to help employees reach the desired output, both in terms of levels output and quality and in terms of the best practice in reaching that output (Secchi and Camuffo, 2016). They are not seen as strict instructions from which deviations are not acceptable but as systematic descriptions of value-added and non-value-added activities that enable employees to perform continuous improvement (Adler and Borys, 1996; Kristensen and Israelsen, 2014). In fact, without standard operating procedures, continuous improvement becomes impossible, as any improvement will be just another variation of the work processes (Liker, 2004). Standard operating procedures are updated to incorporate proven improvements, or they are changed in response to changes in demand or other contingencies (Ahrens and Chapman, 2004). For example, a cell may optimize standard operating procedures affecting other production cells, or changes in market conditions may require manufacturing cells to perform activities differently to meet customer demand.

Standard operating procedures go hand in hand with social control mechanisms, described in Section 2.1. For example, standard operating procedures are visualized (a social visual control mechanism) to employees: pictures of the assembly of parts are made visible on boards in a manufacturing cell, floor markings indicate the flow of materials and finished goods (Kennedy and Widener, 2008), and visual controls indicate whether or not work-in-progress levels are under control (Kristensen and Israelsen, 2014).
Furthermore, employees in lean manufacturing companies undergo training (a social cultural control mechanism) that enables them to understand, perform, and challenge the standard operating procedures (Liker, 2004).

Standard operating procedures work together with non-financial output control mechanisms as well. For example, whiteboards are used in the manufacturing cells to post numbers showing the ability to deliver on time, indicating how well employees are performing. This operating information is used in concert with standard operating procedures to help employees determine whether corrective actions are needed (Kristensen and Israelsen, 2014). The corrective action may adjust current activities, but it may also involve changing and improving the standard operating procedure.

Behavioral controls go beyond standard operating procedures. For example, the Kanban system ensures the replenishment of materials (Shah and Ward, 2007). It includes paper cards that are utilized to pull the right materials to the right places, in the quantities needed, when needed (Emiliani et al., 2003). This demands standards for quantities, materials, procedures for internal customers, and the exact point for when to pull additional materials. One-piece flow and the use of line balancing and level schedules (heijunka) are behavioral controls as well. Optimally, one-piece flow ensures that a part moves to the next operation only when the prior operation is successfully completed (Emiliani et al., 2003). In essence, one-piece flow is then a rule that demands that products are produced only as needed; for this to happen, companies need standard operating procedures that document the sequence of operator work, machine work, and operator movement that is required to produce one unit of a product or part (Miltenberg, 2001). Likewise, line balancing and level schedules demand close relationships with suppliers (Chavez et al., 2015) and standards for production planning and the delivery of products in order to reduce fluctuations in demand and output (Liker, 2004).

2.3 Output control mechanisms

Output control mechanisms consist of performance measurement systems (Kennedy and Widener, 2008). Lean manufacturing companies use detailed non-financial performance measurements to facilitate real-time analyses of cell performance (Fullerton et al., 2014). These measurements track different kinds of cell performance, such as day-by-the-hour, first time through, work-in-progress to standard work-in-progress, and operational equipment effectiveness (Maskell et al., 2012), and they provided fast feedback when problems arise (Barker et al., 1993). These measurements also include past, current, and desired performances, which are supposed to function as motivators for employees and to direct attention to issues that need to be solved. Although different non-financial performance measures are used, this applies for value streams and the facility as well (Emiliani et al., 2003; Maskell et al., 2012). Besides tracking performance and providing feedback, the main purposes of these non-financial performance measures are to align behavior with lean manufacturing objectives (Liker, 2004). This is done in close relationship with social visual control mechanisms, as non-financial performance measures are visually displayed throughout the facility. For example, recurring problems are highlighted on visual boards to initiate kaizens (Emiliani et al., 2003) and to enhance peer pressure in teams (a social cultural control mechanism). These non-financial performance measurements work together with the financial performance measurements presented in quarterly and annual reports (Liker, 2004). Financial performance measurements are also necessary to assist managers and employees in stimulating communication, sending signals related to strategic issues, and fostering learning throughout the organization (Henri, 2006). It is important to distinguish between lean non-financial and financial output control mechanisms, as they are inherently different.
2.4 Hypothesis development

The previous sections described lean management control mechanisms and clarified their interrelatedness. We expect that this interrelatedness will cause complementary effects on firm performance in that the benefits from any lean management control mechanism are greater when the mechanism is accompanied and integrated with the other lean management control mechanisms (Roberts, 2004). For example, performance measurement systems (output control mechanisms) drive behavior to a greater extent and are more likely to direct employees’ attention to problems if they are visualized through social controls. The effect of peer pressure (a social control mechanism) will be higher if boards containing skill matrices are visualized (an output control mechanism) to other employees. Standard operating procedures (a behavioral control mechanism) may be tacit knowledge for employees, but they are more effective if they are visualized, ensuring that all employees work according to the best standard currently known. The visualization of standards also enables employees to challenge and improve these standards. Additionally, the effectiveness of standard operating procedures will likely be higher if all employees are trained according to these standards (social control mechanism).

When complementarities exist among management control mechanisms, a firm needs to coordinate the use of these management control mechanisms by implementing them simultaneously. Thus, we follow the same procedure as Tanriverdi and Venkatraman (2005) and develop a latent second-order construct. The first level of this construct captures the sub-additive effects arising from social, behavioral, and output control mechanisms, and the second level captures the super-additive effects from the complementarity of management control mechanisms. When assessing the performance effects of a complementary system of management control mechanisms, we have to compare the performance effects of individual management control mechanisms with the performance effects of the complementarity among management control mechanisms, and we have to ensure that the complementarity performance effects outweigh the individual effects (Tanriverdi and Venkatraman, 2005; see also Ichniouwski et al., 1997; Whittington et al., 1999). Following Tanriverdi and Venkatraman’s (2005) procedure, we develop two competing hypotheses to test whether the performance effects of management control mechanisms in lean manufacturing companies are contingent on the complementarity of these management control mechanisms or whether the individual management control mechanism has an independent direct effect on performance: a “strong form,” stating that the complementarity of management control mechanisms will have a direct positive effect on firm performance; and a “weak form,” stating that each management control mechanism will have an independent direct positive effect on firm performance:

- **H1** (strong form). The complementarity of social control mechanisms, behavioral control mechanisms, and output control mechanisms has a positive effect on firm performance.
- **H2** (weak form). Social control mechanisms, behavioral control mechanisms, and output control mechanisms have independent positive effects on firm performance.

Figure 1 includes a conceptual model of the complementarity **H1**.

3. Methods

The survey was distributed online to 4,357 subjects, representing 697 manufacturing facilities, in September 2012, and responses were received until December 2012.
The subjects were identified from the Shingo Prize[4], Organization database of individuals who had expressed an interest in receiving information about lean principles, Shingo seminars and workshops, and the Shingo Prize. We received responses from 510 individuals, representing 368 different facilities, yielding a response rate of 11.70 percent which is similar to other research papers on lean manufacturing (e.g. Hofer et al., 2012; Shah and Ward, 2003). We averaged responses from plants from which we received multiple responses, leaving us with a usable sample size of 368 and a facility response rate of 52.8 percent. Collectively, the 368 facilities represented 195 different organizations. In total, 30 percent of the organizations produced vehicles or provided components to the automotive industry, 29 percent produced healthcare-related products, 23 percent made products for the aerospace industry, and 19 percent produced components for the department of defense.

Of the facilities, 52 percent had more than 500 employees and 53 percent of the facilities had sales of over $100 M. The average management experience of the respondents within their current firms was 11.3 years. This is important to our study, as experienced managers are likely to understand our holistic set of questions regarding management control, lean manufacturing, and performance in their facilities[5]. Of the respondents, 53.5 percent were responsible for lean, quality, or continuous improvement. Survey questions were intended to assess the level of lean manufacturing and management control implementation at the respondents’ facilities as well as to obtain a self-assessment of firm performance.

In the following sections, we describe how we developed our variables. We also go through our statistical tests and explain why we decided to utilize Tanriverdi and Venkatraman’s (2005) test for complementarity. Figure 2 illustrates the sequence of the statistical tests.

3.1 Measures
Although the questionnaire included 148 questions, we only included a portion for analysis in the present paper. We drew upon Kennedy and Widener (2008) in developing most management control mechanism items, and we adapted several items from Fullerton et al. (2013, 2014).
We developed four items covering cultural social control mechanisms, intended to cover the degree to which the entire facility is trained in lean principles (CLTR 4), employee empowerment (CLTR 3 and CLTR 1), and peer pressure (CLTR 8). Furthermore, we developed three additional items, CLTR 5, CLTR 6, and CLTR 7, intended to capture the degree to which the facilities work with continuous improvement, the degree to which management is focused on eliminating waste, and the degree to which lean thinking has permeated all operations, respectively. CLTR 2 was adapted from Fullerton et al. (2013) and was intended to cover the degree to which management is committed to quality-related training. Of the seven items covering visual social control mechanisms, MAS 2, MAS 4, MAS 5, and MAS 7 were adapted from Fullerton et al. (2013), while the remaining three items were developed in accordance with Kennedy and Widener (2008). All items were intended to capture the degrees of different types of visualization.

Three of four items covering behavioral control mechanisms were adapted from Fullerton et al. (2013) and were intended to cover the degree of facilities’ use of standardization of manufacturing procedures (MFG 1), a Kanban system (MFG 2), and one-piece flow (MFG 3), and we developed MFG 4 to capture the use of line balancing and level schedules.

The three items covering non-financial output controls were intended to capture the importance of non-financial performance measures related to cell performance (PRF 1), value stream performance (PRF 2), and facility performance (PRF 3). As these measures are rather generic, we follow the same procedure as Fullerton et al. (2013) and include a test for criterion validity where we correlate our non-financial output controls with criterion variables in order to demonstrate plausibility. This test can be found in Table AI. We developed four additional items covering financial output control mechanisms, intended to capture the importance of performance measures related to market share (PRF 4), cash flow (PRF 5), overall financial results (PRF 6), and customer satisfaction (PRF 7). One of the six items covering performance (LIMP 3) was adapted from Fullerton et al. (2014), while we developed the remaining items in order to cover the extent to which lean initiatives have freed inventory resources (LIMP 1), improved capacity management effectiveness (LIMP 2), improved quality (LIMP 4), improved communication (LIMP 5), reduced costs (LIMP 6), and improved profitability (LIMP 7). Thus, our performance items cover both a goal-centered and an accounting approach (Kihn, 2005). Survey items can be found in Table AII.

All items were measured on a five-point labeled Likert scale. Eustler and Lang (2015) have shown that labeled scales are superior to unlabeled scales as they reduce measurement error and response bias.

3.2 Exploratory factor analysis
We conducted an exploratory factor analysis including our exogenous variables with oblique rotation. We removed one item that loaded greater than 0.4 on more than one variable[6]. After the removal of one item, we conducted another exploratory factor analysis, which yielded five factors with eigenvalues greater than 1, collectively explaining 66.8 percent of the variance: cultural social controls, visual social controls, behavioral controls, non-financial output controls, and financial output controls. Additionally, we performed an exploratory factor analysis for the performance items yielding one factor with an eigenvalue greater than 1, explaining 65.5 percent of the variance. Along with the
exogenous factors, the performance factor represents the variables used in this study (see Table II). All factors’ Cronbach’s α’s are between 0.786 and 0.913 (see Table III), demonstrating good to excellent reliability (Kline, 2011).

### 3.3 Confirmatory factor analysis

We perform a confirmatory factor analysis in AMOS 23 including our factors, using maximum likelihood estimation. This is a two-step procedure where the measurement model without structural paths is evaluated to ensure that it fits, and this is followed by an evaluation of the entire structural model (Hair et al., 2014). We evaluate the measurement model using several fit indices, as recommended by Kline (2011). We assess $\chi^2$ to degrees of freedom (Bollen, 1989), as it seems to be the consensus in the SEM literature, although Kline (2011) stated that there is little statistical and logical foundation for using this measure of model fit. We assess the root mean square error of approximation (RMSEA) and the

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Cultural social controls</th>
<th>Visual social controls</th>
<th>Non-fin. output controls</th>
<th>Financial output controls</th>
<th>Behavioral controls</th>
<th>Firm performance</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLTR1</td>
<td>0.585</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.49</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>CLTR2</td>
<td>0.517</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.84</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>CLTR3</td>
<td>0.707</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.33</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>CLTR4</td>
<td>0.568</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.33</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>CLTR5</td>
<td>0.727</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.42</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>CLTR6</td>
<td>0.851</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.44</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>CLTR7</td>
<td>0.755</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.50</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>CLTR8</td>
<td>0.766</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.18</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>MAS1</td>
<td>-0.629</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.68</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>MAS2</td>
<td>-0.651</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.08</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>MAS3</td>
<td>-0.724</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.35</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>MAS4</td>
<td>-0.849</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.72</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>MAS5</td>
<td>-0.745</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.39</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>MAS6</td>
<td>-0.854</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.67</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>MAS7</td>
<td>-0.745</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.37</td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td>PRF1</td>
<td>-0.851</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.18</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>PRF2</td>
<td>-0.790</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.02</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>PRF3</td>
<td>-0.852</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.18</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>PRF4</td>
<td>0.825</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.43</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>PRF5</td>
<td>0.869</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.72</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>PRF6</td>
<td>0.749</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.17</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>PRF7</td>
<td>0.509</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.19</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>MFG1</td>
<td>-0.782</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.86</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>MFG2</td>
<td>-0.727</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.52</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>MFG3</td>
<td>-0.719</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.27</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td>MFG4</td>
<td>-0.779</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.50</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>LIMP1</td>
<td>0.754</td>
<td></td>
<td>0.754</td>
<td>3.24</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP2</td>
<td>0.850</td>
<td></td>
<td>3.59</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP3</td>
<td>0.848</td>
<td></td>
<td>3.72</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP4</td>
<td>0.785</td>
<td></td>
<td>3.64</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP5</td>
<td>0.768</td>
<td></td>
<td>3.62</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP6</td>
<td>0.828</td>
<td></td>
<td>3.60</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP7</td>
<td>0.829</td>
<td></td>
<td>3.53</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** KMO of sampling adequacy for the management control mechanism factors: 0.944, Bartlett’s test of sphericity is significant $p < 0.000$. KMO of sampling adequacy for the firm performance factor 0.887, Bartlett’s test of sphericity is significant $p < 0.000$. The KMO values above 0.5 and the significance of the Bartlett’s test of sphericity indicates that the data are suitable for exploratory factor analysis, and that there are patterns among items (Field, 2005). Only loadings exceeding 0.400 are shown.
standardized root mean square residual (SRMR). Additionally, we evaluate the comparative fit index (CFI) (Bentler, 1990), incremental fit index (IFI) (Bollen, 1989), and Tucker-Lewis index (TLI) (Tucker and Lewis, 1973). In general, there are no accepted minimal thresholds for what constitutes acceptable model fit (Schermelleh-Engel et al., 2003). However, there are suggested parameters in published academic work for what would represent acceptable fit: $\chi^2$ to degrees of freedom should be less than 3, indicating acceptable fit (Kline, 2005); a RMSEA value below 0.08 would indicate acceptable fit (Browne and Cudeck, 1993; Kline, 2011); a SRMR value below 0.1 indicates acceptable fit (Schermelleh-Engel et al., 2003); and CFI, IFI, and TLI are evaluated for their closeness to 1.0 (Byrne, 2010) with values over 0.9 (Bentler, 1992; Kline, 2005), indicating acceptable fit. Finally, we evaluate the Consistent Akaike’s Information Criterion, addressing the issue of parsimony in the assessment of model

<table>
<thead>
<tr>
<th>Factor indicators</th>
<th>Standardized loadings</th>
<th>$t$-value (all significant $p &lt; 0.01$)</th>
<th>CR</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultural social controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTR1</td>
<td>0.71</td>
<td>a</td>
<td>0.908</td>
<td>0.904</td>
</tr>
<tr>
<td>CLTR2</td>
<td>0.62</td>
<td>11.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTR3</td>
<td>0.71</td>
<td>13.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTR4</td>
<td>0.50</td>
<td>10.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTR5</td>
<td>0.76</td>
<td>14.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTR6</td>
<td>0.82</td>
<td>15.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTR7</td>
<td>0.83</td>
<td>15.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLTR8</td>
<td>0.88</td>
<td>16.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual social controls</td>
<td></td>
<td></td>
<td>0.912</td>
<td>0.909</td>
</tr>
<tr>
<td>MAS1</td>
<td>0.70</td>
<td>13.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAS2</td>
<td>0.78</td>
<td>15.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAS3</td>
<td>0.73</td>
<td>14.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAS4</td>
<td>0.83</td>
<td>16.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAS5</td>
<td>0.80</td>
<td>15.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAS6</td>
<td>0.81</td>
<td>16.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAS7</td>
<td>0.76</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior controls</td>
<td></td>
<td></td>
<td>0.826</td>
<td>0.821</td>
</tr>
<tr>
<td>MFG1</td>
<td>0.65</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFG2</td>
<td>0.75</td>
<td>11.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFG3</td>
<td>0.78</td>
<td>12.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFG4</td>
<td>0.77</td>
<td>12.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-financial output controls</td>
<td></td>
<td></td>
<td>0.913</td>
<td>0.913</td>
</tr>
<tr>
<td>PRF1</td>
<td>0.89</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRF2</td>
<td>0.90</td>
<td>24.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRF3</td>
<td>0.86</td>
<td>22.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial output controls</td>
<td></td>
<td></td>
<td>0.805</td>
<td>0.797</td>
</tr>
<tr>
<td>PRF4</td>
<td>0.75</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRF5</td>
<td>0.77</td>
<td>12.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRF6</td>
<td>0.70</td>
<td>12.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRF7</td>
<td>0.62</td>
<td>10.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm performance</td>
<td></td>
<td></td>
<td>0.913</td>
<td>0.912</td>
</tr>
<tr>
<td>LIMP1</td>
<td>0.72</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP2</td>
<td>0.82</td>
<td>15.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP3</td>
<td>0.83</td>
<td>15.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP4</td>
<td>0.75</td>
<td>14.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP5</td>
<td>0.79</td>
<td>14.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP6</td>
<td>0.78</td>
<td>14.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIMP7</td>
<td>0.72</td>
<td>13.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: $\chi^2$ to degrees of freedom: 2.299, RMSEA: 0.060, SRMR: 0.054, IFI: 0.923, TLI: 0.915, CFI: 0.922, CAIC: 0.429 (1,663.439/3,875.435 saturated model). “a” Indicates a loading fixed to 1.

Table III.
Confirmatory factor analysis, composite reliability, and Cronbach’s $\alpha$
fit, taking sample size into account (Bozdogan, 1987), where the ratio of the hypothesized model and the saturated model should be less than 1 (Byrne, 2010). Although the $\chi^2$ is significant ($p < 0.001$), the $\chi^2$ to degrees of freedom is less than 3, and fit indices are more than acceptable (see Table III).

To assess construct validity, we investigate the factors’ convergent validity, composite reliability (CR), and discriminant validity. All our factors show good convergent validity, as their average variance extracted (AVE) is above 0.5 (see Table IV) and their CR is well above 0.7 (Hair et al., 2014). Furthermore, as indicated in Table III, all factor loadings (standardized coefficients) are above 0.5 (Bagozzi and Yi, 1988). Discriminant validity is assessed by comparing the square root of the AVE of the factors with their correlation (Fornell and Larcker, 1981), where the square root AVE of individual factors should be greater than the interfactor correlation. Square root AVE of factors is indicated at the diagonal of Table IV and is greater than the interfactor correlations[7]. Additionally, none of the interfactor correlations exceed their $\alpha$, which is another indicator of discriminant validity (Crocker and Algina, 1986). Table IV also indicates that all factors correlated significantly. Our measurement model did not indicate multicollinearity issues, as none of the variance inflation factors exceeded 2.8, and all tolerance statistics exceeded 0.36.

Before running the two full structural models, we also test all relationships from exogenous variables to performance for linearity. All relationships are significantly linear $p < 0.01$ and have $R^2$ values ranging from 0.146 to 0.656 and $F$-values between 62.658 and 697.191. In addition, the number of free parameters to be estimated compared with the sample size is well above the minimum ratio of 1:5 recommended by Worthington and Whittaker (2006) in both the first-order structural model and the second-order structural model.

### 3.4 Testing for complementarity

There are several strategies when testing for complementarities in research. Ennen and Richter (2010) divided these strategies into two main categories: the interaction approach and the systems approach. The interaction approach is of a reductionist character (Drazin and Van de Ven, 1985), as it only includes pairs of interactions and their main effects in a regression model. This is often a function of statistical necessity, as individual variables in complementary systems are heavily correlated and, furthermore, heavily correlated with the interaction term. When the main variables and their pair-wise interaction terms are heavily correlated, coefficient estimates obtained from the regression model do not reflect the inherent effects of any particular independent variable on the dependent variable but only the marginal effects or the partial effects, given the other, independent variables in the model (Tanriverdi and Venkatraman, 2005). Our independent variables are significantly correlated, as shown in Table IV. Likewise, our multiplicative interaction terms are heavily correlated with each other and with their main variables (correlations ranging from 0.311 $p < 0.001$ to 0.935 $p < 0.001$)[8]. Furthermore, by focusing only on pairs of interactions,

<table>
<thead>
<tr>
<th>Factor</th>
<th>No. of measures</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-financial output controls</td>
<td>3</td>
<td>0.883</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.779</td>
</tr>
<tr>
<td>Visual social controls</td>
<td>7</td>
<td>0.620**</td>
<td>0.773</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.598</td>
</tr>
<tr>
<td>Cultural social controls</td>
<td>8</td>
<td>0.667**</td>
<td>0.678**</td>
<td>0.739</td>
<td></td>
<td></td>
<td></td>
<td>0.547</td>
</tr>
<tr>
<td>Behavioral controls</td>
<td>4</td>
<td>0.528**</td>
<td>0.607**</td>
<td>0.609**</td>
<td>0.738</td>
<td></td>
<td></td>
<td>0.544</td>
</tr>
<tr>
<td>Financial output controls</td>
<td>4</td>
<td>0.411**</td>
<td>0.447**</td>
<td>0.428**</td>
<td>0.475**</td>
<td>0.713</td>
<td></td>
<td>0.508</td>
</tr>
<tr>
<td>Firm performance</td>
<td>7</td>
<td>0.627**</td>
<td>0.674**</td>
<td>0.762**</td>
<td>0.718**</td>
<td>0.491**</td>
<td>0.774</td>
<td>0.600</td>
</tr>
</tbody>
</table>

**Notes:** All measures are a labeled Likert scale from 1 to 5. Square roots of AVE are shown at the diagonal. **Significant at the $p < 0.01$ level

Table IV. Factor correlations, squared average variance extracted, and average variance extracted.

Complementary management control mechanisms
researchers that are not able to detect the expected complementarity between two variables might overlook that the expected complementarity is a function of a third variable (Ennen and Richter, 2010). Our theory concerns complementarities among multiple variables. Given the theoretical development and explanations leading to our complementarity hypothesis, the interpretational problems inherent in the interaction approach render it an ineffective means of testing the hypothesis.

The systems strategy testing complementarity involves focusing on a holistic set of variables (Ennen and Richter, 2010). However, Ennen and Richter (2010) do not elaborate on the statistical testing techniques of this strategy. Profile deviation analysis is suggested by Gerdin and Greve (2004). Studies that use profile deviation analysis segment data based on a criterion variable and find the ideal state of systems within each of these segments (see e.g. Hult et al., 2007). As a second step, researchers use the city block distance or the Euclidian distance, expecting that the deviations from the ideal state are negatively associated with performance. However, the city-block distance only accounts for additive effects, and it is unclear exactly what is captured by the Euclidian distance. Another possibility when pursuing systems strategy is to apply higher-order interactions in a regression model. However, this approach will increase the correlations between individual variables and their multiplicative interactions, leading to interpretational problems of the regression model (Tanriverdi and Venkatraman, 2005). Other studies that apply the systems strategy attempt to capture the nature of organizational systems by using a categorical variable that studies whether or not a particular factor is in place (e.g. Furlan et al., 2011). However, this approach provides little information on the nature of the relationships that drive the complementarity effects observed (Ennen and Richter, 2010).

As the tests described here were not appropriate for testing our hypotheses on complementarity, we sought an alternative statistical method and decided to utilize the approach applied by Tanriverdi and Venkatraman (2005). Tanriverdi and Venkatraman (2005) constructed two models in order to test for complementarity: a first-order model to capture the sub-additive effects of their variables on performance and a second-order factor model to account for the multilateral interactions and covariance among their variables, in order to test for the complementary effects on performance. A second-order factor is an entity that is reflected by first-order factors serving as its indicators (Williams et al., 2004) and is the main source of covariance among first-order factors; it explains why the first-order factors coexist and co-vary with each other (Rindskopf and Rose, 1988). Utilizing Tanriverdi and Venkatraman’s (2005) procedure, we avoid the interpretational challenges of the other tests for complementarity of multiple variables (in our case, control mechanisms), and we can compare the additive effects on firm performance with the complementary effects on firm performance. We are thus able to test both our hypotheses and to determine whether the complementary effects outweigh the additive effects as well as whether some of the management control mechanisms affect firm performance in isolation.

3.5 Assessment of the second-order measurement model
Following Tanriverdi and Venkatraman’s (2005) procedure[9], we need to compare the first-order measurement model where we correlate our management control mechanisms with the second-order measurement model in order to assess the existence of a second-order model and to ensure the multidimensionality, construct, and convergent validity of the second-order model. Marsh and Hocevar (1985) developed the target coefficient statistic, which is the ratio of the χ² of the first-order model to the χ² of the second-order model. The target coefficient has an upper limit of 1.0 (Tanriverdi and Venkatraman, 2005), and support for the existence of a second-order factor becomes stronger when the target coefficient approaches unity (Marsh and Hocevar, 1985). The value of the target coefficient of our second-order complementarity factor is 0.98, indicating that a second-order factor
explains 98 percent of the relations among the first-order factors. Furthermore, all second-order factor loadings are highly significant ($p < 0.001$), providing further acceptance of a second-order model. Collectively, these results support the existence, multidimensionality, convergent and discriminant validity, and reliability of a second-order complementarity construct (Tanriverdi and Venkatraman, 2005) (see Table V).

4. Empirical tests and results

4.1 Test of hypotheses

The figures depict the models of our two competing hypotheses. Figure 3 shows a graphical representation of the model for testing $H1$. This depicts our management control mechanisms modeled initially as first-order factors. The second-order factor in the figure models the complementarity among our management control mechanisms by accounting for their covariance and multilateral interactions, and the directions of the structural links are from the second-order factor to the first-order factor, indicating that all the management control mechanisms are adapted simultaneously and systematically. In order to test our hypothesis, the second-order factor is related to firm performance. Figure 4 shows a graphical representation for testing $H2$. It shows the management control mechanisms as first-order factors, models their pair-wise covariance, and relates the management control factors additively to firm performance. In Figure 3, the structural parameter from the complementarity second-order factor to firm performance is positive and significant (standardized $\beta$ coefficient: 0.927, $p < 0.001$, $R^2$: 0.859), providing support for $H1$, the strong form. This finding indicates that a second-order factor accounting for the complementarity among management control mechanisms has a positive effect on firm performance. In Figure 4, only two of the five structural parameters, cultural social control mechanisms (standardized $\beta$ coefficient: 0.399 $p < 0.001$) and behavioral control mechanisms (standardized $\beta$ coefficient: 0.400 $p < 0.001$, collective $R^2$ from all additive effects: 0.805), from management control mechanisms to firm performance are significant (also see Table VI, Panel A). Financial and non-financial output control mechanisms and

Panel A: fit indices for the first-order measurement model and the second-order measurement model

<table>
<thead>
<tr>
<th>Fit indices</th>
<th>First-order measurement model</th>
<th>Second-order measurement model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>605.899</td>
<td>617.845</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>289</td>
<td>294</td>
</tr>
<tr>
<td>$\chi^2$ to degrees of freedom</td>
<td>2.097</td>
<td>2.012</td>
</tr>
<tr>
<td>IFI</td>
<td>0.946</td>
<td>0.938</td>
</tr>
<tr>
<td>TLI</td>
<td>0.939</td>
<td>0.938</td>
</tr>
<tr>
<td>CFI</td>
<td>0.946</td>
<td>0.944</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.055</td>
<td>0.055</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.056</td>
<td>0.058</td>
</tr>
<tr>
<td>CAIC (default model to saturated model)</td>
<td>0.422</td>
<td>0.422</td>
</tr>
<tr>
<td>Target statistic: 0.980 (605.899/617.845)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: first-order factor loadings on complementary factor

<table>
<thead>
<tr>
<th>Relationships</th>
<th>Standardized coefficient</th>
<th>$t$-values (all significant at $p &lt; 0.001$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-financial output controls ↔ complementarity factor</td>
<td>0.806</td>
<td>9.815</td>
</tr>
<tr>
<td>Visual social controls ↔ complementarity factor</td>
<td>0.844</td>
<td>9.060</td>
</tr>
<tr>
<td>Cultural social controls ↔ complementarity factor</td>
<td>0.866</td>
<td>10.101</td>
</tr>
<tr>
<td>Behavioral controls ↔ complementarity factor</td>
<td>0.793</td>
<td>a</td>
</tr>
<tr>
<td>Financial output controls ↔ complementarity factor</td>
<td>0.549</td>
<td>7.205</td>
</tr>
</tbody>
</table>

Note: “a” Indicates a loading fixed to 1

Table V. Assessment of the first-order and second-order measurement models
social visual control mechanisms do not contribute to performance in isolation. Thus, $H2$, the weak form, is not supported. Both the standardized $\beta$ coefficient and $R^2$ from the complementary factor to firm performance are greater than the collective $R^2$ and the standardized $\beta$ coefficients in the additive model. These results suggest that the complementary effects on firm performance among the complete set of lean management control mechanisms outweigh their individual performance effects, providing further acceptance for[10].

We have decided to report the fit indices in Figures 3 and 4, for which there are consensus in the structural equation modeling literature (Kline, 2011), although Tanriverdi and Venkatraman (2005) chose not to do so. All fit indices indicate acceptable fit.

Notes: The bold arrow indicates a significant relationship from the exogenous variable to the endogenous variable. $^a$Visual social controls; $^b$cultural social controls; $^c$non-financial output controls; $^d$financial output controls; $^e$behavioral controls.

$\chi^2 = 1,136.651$; degrees of freedom = 489; $\chi^2$ to degrees of freedom = 2.324; IFI = 0.92; TLI = 0.913; CFI = 0.919; RMSEA = 0.060; SRMR = 0.056; CAIC (default model to saturated model) = 0.042.

*Significant at $p < 0.001$
5. Discussion and conclusion

This study focused on complementarities among management control mechanisms in lean manufacturing companies. Little research has been carried out on this topic, which is rather paradoxical, as lean manufacturing is recognized as an enterprise-wide system consisting of interdependent practices (Liker, 2004; Maskell et al., 2012). Our aim with this research was to study lean management control mechanisms and their complementary effects on firm performance. Earlier research provides limited evidence of complementarity among lean management control mechanisms. Emiliani et al. (2003) and Kennedy and Widener (2008) were single firm studies and found that lean management control mechanisms were interrelated, but did not provide evidence of complementary effects from lean management control mechanisms to firm performance. Kristensen and Israelsen (2014) was a single firm study showing that greater balance among management control mechanisms led to greater

Notes: Bold arrows indicate significant relationships from exogenous variables to the endogenous variable. *Visual social controls; b cultural social controls; c non-financial output controls; d financial output controls; e behavioral controls. \( \chi^2 = 1,103.884 \); degrees of freedom = 480; \( \chi^2 \) to degrees of freedom = 2.300; IFI = 0.923; TLI = 0.915; CFI = 0.922; RMSEA = 0.060; SRMR = 0.054; CAIC (default model to saturated model) = 0.043. *Significant at \( p < 0.001 \)

Figure 4. Hypothesis test 2 (weak form)
firm performance, but their method made it difficult to capture the covariance and interactions among lean management control mechanisms. Fullerton et al. (2013) was a cross-sectional study and found that management control mechanisms were interrelated. However, the study did not provide evidence of the complementary effects from lean management control mechanisms to firm performance, and did not encompass the complete set of management control mechanisms.

Informed by the lean manufacturing literature and complementary theory, we expected that lean management control mechanisms were complementary. We utilized the holistic framework developed by Kennedy and Widener (2008), which characterizes lean management control mechanisms as social, behavioral, and output control mechanisms. In order to confirm that management control mechanisms were complementary, we constructed two competing hypotheses. The first hypothesis predicted that the complementarity of management control mechanism was positively related to firm performance. The second hypothesis predicted that the management control mechanisms were independently, additively related to firm performance. By constructing two competing hypotheses, we were able to compare the performance effects of individual system components with the performance effects of the complementarity among system components, and we were able to point out the conditionality of individual effects on the effects of other system components (Tanriverdi and Venkatraman, 2005).

We contribute to the literature on lean management control mechanisms in two major ways. We are the first to show that the complementary effects among lean management control mechanisms outweigh their additive effects on firm performance. Thus, firm performance will suffer as a result of implementations that do not consider the complementarity among management control mechanisms (Roberts, 2004). Furthermore, only social cultural control mechanisms and behavioral control mechanisms were independently related to firm performance. Second, this research adds cross-sectional empirical evidence that the full set of lean management control mechanisms is complementary. We also add greater granularity to the understanding of lean management control mechanisms because we distinguish financial output controls from non-financial controls as well as social visual controls from cultural visual controls, and we add a detailed analysis of their systematic interrelatedness. In other words, we provide evidence of five different management control mechanisms compared with the three found in Kennedy and Widener’s (2008), and Kristensen and Israelsen’s (2014) studies. Inspired by Ouchi (1979), we also extend the Kennedy and Widener’s (2008) framework by incorporating lean thinking into social control mechanisms. The greater granularity and greater level of detail are important steps forward in understanding lean management control mechanisms.

<table>
<thead>
<tr>
<th>Panel A: hypotheses tests (weak form)</th>
<th>Independent variable</th>
<th>Dependent variable</th>
<th>Standardized coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSC</td>
<td>→</td>
<td>Firm performance</td>
<td>0.094</td>
</tr>
<tr>
<td>CSC</td>
<td>→</td>
<td>Firm performance</td>
<td>0.399***</td>
</tr>
<tr>
<td>OUTNF</td>
<td>→</td>
<td>Firm performance</td>
<td>0.058</td>
</tr>
<tr>
<td>OUTF</td>
<td>→</td>
<td>Firm performance</td>
<td>0.082</td>
</tr>
<tr>
<td>EC</td>
<td>→</td>
<td>Firm performance</td>
<td>0.400***</td>
</tr>
<tr>
<td>R² firm performance: 0.805</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: hypothesis test (strong form)</th>
<th>Independent variable</th>
<th>Dependent variable</th>
<th>Standardized coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complementarity</td>
<td>→</td>
<td>Firm performance</td>
<td>0.927***</td>
</tr>
<tr>
<td>R² firm performance: 0.859</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table VI. Results

**Note:** ***Significant at the p < 0.001 level
To illustrate our findings, consider that non-financial output control mechanisms are not recognized as complementary with peer pressure (a social cultural control mechanism) in the system. That will lead to a reduction of the motivational effects otherwise promoted by non-financial output control mechanisms. Likewise, the effects of structuring the manufacturing facility with high visibility (a social visual control mechanism) are reduced if managers do not recognize the complementarity with training in lean principles (a social cultural control mechanism), as employees will not be able to assist other manufacturing cells in preventing problems or improving work processes. Furthermore, if managers do not recognize that visualization of quality data (a social visual control mechanism) is complementary with standardization (a behavioral control mechanism), the effects of visualization of quality data are reduced, as it is difficult to leverage for continuous improvement, because employees have no baseline from which they can test potential improvements. The performance effects of financial and non-financial output control mechanisms and of social visual control mechanisms are thus not isolated additive effects; they affect performance through their complementarity with social cultural control mechanisms and behavioral control mechanisms.

In a lean manufacturing milieu, social cultural control mechanisms and behavioral control mechanisms are then not only enhancers of firm performance, but also enablers for the performance effects of financial and non-financial output control and social visual control mechanisms. In a similar vein, the effects of social cultural control mechanisms and behavioral control mechanisms on firm performance are greater when they are accompanied by non-financial output control and social visual control mechanisms in a complementary system. This underlines that the greatest benefits from lean management control mechanisms arise when they are implemented in a complete, systematic manner.

Methodologically, this study makes two contributions to the management control literature. First, we use a second-order factor technique to find evidence of complementarity among management control mechanisms. This technique is new to this body of literature and it overcomes the struggles of other techniques testing for complementarities. The second methodological advance of this study is that we show the specifics of management control mechanisms in a lean manufacturing context and show how individual management control mechanisms are related (Malmi and Brown, 2008).

Our findings have important managerial implications. First, companies will not achieve the full performance potential of implementing lean manufacturing if they decide to employ a system where some of the management control mechanisms are missing. In line with this reasoning, and if a company has already employed for example non-financial output control mechanisms, it should invest in implementing the remaining management control mechanisms rather than putting more effort into the existing one. Second, the implementation of all lean management control mechanisms affects the entire company, and employees might have to unlearn old principles and practices before new ones can be put fruitfully into use. Thus, the implementation of the full set of management control mechanisms should be performed with a great emphasis on company-wide coordination, and companies would benefit from preparing employees thoroughly before embarking on the lean manufacturing journey. Third, it is important for decision makers to understand that the performance effect of the implementation of one management control mechanism is dependent on the level of implementation of another management control mechanism, and vice versa, and that the company will not obtain the full performance effects until the system of management control mechanisms is completely implemented. Therefore, although initial performance effects might be lower than expected, the company should not hesitate with respect to increasing the level of the implementation of lean management control mechanisms. The implementation of the full set of complementary lean management control mechanisms requires complex organizational change and coordination, but this research enables decision makers a greater prior understanding of how the
management control mechanisms work together. Thus, our research can assist and guide
decision makers in overcoming some hesitations related to the implementation, and leaves less
to understand after the implementation. Fourth, the set of questionnaire items that we
developed in this research can be applied by practitioners during lean audits to ensure that
they are on track and reaching lean manufacturing objectives, and the set of items can be used
as a benchmarking tool between business units. Fifth, our evidence suggests that decision
makers should understand that financial output control mechanisms remain important in lean
manufacturing companies. In the literature, e.g. Johnson (1992), it is typically noted that such
control mechanisms should be avoided and substituted with non-financial control mechanisms,
but we have shown that non-financial and financial control mechanisms are complementary.
Finally, lean management control mechanisms might be relatively easy to replicate between
companies. Furthermore, knowledge of lean principles and practices is wide-spread. After all,
these principles and practices have received abundant attention since the late 1980s
(Bhamu and Sangwan, 2014). Therefore, despite that initial costs might be high, companies
should go far in order to understanding the complementarity among the complete set of
management control mechanisms as it may lead to a sustainable competitive advantage
because it is difficult for competitors to replicate (Porter, 1996).

6. Future research and limitations
As with other studies, this study has its limitations. As our study is of a cross-sectional
nature, it is difficult to claim causal inferences, and we cannot rule out that unobserved
factors may be driving our evidence. Rather, our evidence must be considered as consistent
with our theoretical arguments. Furthermore, our sample is not random, as it was drawn
from a population of lean companies. This reduces the generalizability of our evidence to
other manufacturing regimes, but it also increases the likelihood of the population
understanding the survey questions and consequently helps alleviate some of the concerns
about data collection in survey research (Fullerton et al., 2013). Finally, surveying only one
respondent in each firm represents a potential common method bias problem. However,
we addressed this limitation and found that it was not a concern.

Our study suggests that examining the benefits or effects of financial and non-financial
control mechanisms and social visual control mechanisms in isolation at lean companies
may lead to inconsistent results due to a failure to control for social cultural and behavioral
control mechanisms. Future research on management control in lean companies must then
encompass a focus on the entire set of management control mechanisms. The simultaneous,
systematic implementation of lean management control mechanisms might overwhelm
employees’ absorptive capacity (Cohen and Levinthal, 1990). A possible future research
endeavor is then to clarify if the effects of lean management control mechanisms on firm
performance are affected by the length of time companies have used lean manufacturing.
A second future research idea is to clarify whether our findings are applicable to more
loosely coupled manufacturing regimes. In these manufacturing regimes, the individual
management control mechanism might work, as practices are less interdependent
(Roberts, 2004). Testing for complementarities among management controls has recently
been debated (see Grabner and Moers, 2013). We consider the second-order technique as an
important addition to this debate, and we suggest that future management control research
on complementarities should consider using the second-order technique.

Notes
1. We use the label “management control mechanisms” as Kennedy and Widener (2008) used this
label. We believe that it is equivalent to the label “management control forms” used in other
studies, e.g., Kristensen and Israelsen (2014).
2. Management control is defined by Anthony (1965, p. 17) as, “the process by which managers ensure that resources are obtained and used effectively and efficiently in the accomplishment of the organization’s objectives.”

3. Ouchi (1979, pp. 837) stated that some of the characteristics of clan controls are to ensure that employees try to achieve the “right” objectives.

4. The Shingo Prize is an award given to companies based on their world-class results and organizational culture. The database includes many companies, as most organizations do not wait to challenge for the Shingo Prize until they are likely to win it.

5. As the large majority of respondents had management experience and were responsible for lean at their facility, our constructs might be subject to common method bias. To reduce these concerns, we perform a Harman’s one factor test including all our latent variables. There is a potential bias if the majority of the variance is explained by one factor (Podsakoff and Organ, 1986). The test shows that the concern for common method bias is low, as a one factor solution only accounts for 45 percent of the total variance.

6. The 0.4 cut-off have been used in prior research on lean manufacturing (e.g., Fullerton and Wempe, 2009; Fullerton et al., 2014). The removal of one item did not affect the composition of the five factors.

7. Squared AVE to inter-factor correlations is computed in SPSS 23. We compared the squared AVE to the inter-factor correlations in AMOS 23 as well. This test revealed discriminant validity issues only concerning the performance factor, the social controls one factor, and the behavioral controls factor. All of our factors correlated less than 0.85, not indicating poor discriminant validity (Kenny, 2012). Kenny (2012) also suggested restricting the correlation between two factors to 1, which is similar to collapsing the two factors (Hair et al., 2014). This is done to investigate if a one-factor model is more appropriate than a two-factor model. A two-factor model is appropriate if $\chi^2/df_{\text{diff.}}$ is significant (Hair et al., 2014). We performed a test in AMOS 23 where we constrained correlations between both the performance factor and behavioral controls and the performance factor and social controls 1. In both instances, a two-factor model fitted the data significantly better: restricting the correlation to one between performance and behavioral controls yields a $\chi^2$ of 1,199.22 and degrees of freedom: 482, resulting in a significant $\chi^2/df_{\text{diff.}}$ ($p < 0.01$) and the following fit indices: RMSEA: 0.064, SRMR: 0.1307, IFI: 0.911, TLI: 0.902, and CFI: 0.911. Restricting the correlation to one between performance and social controls 1, on the other hand, yields a $\chi^2$ of 1,143.806 and degrees of freedom: 482, resulting in a significant $\chi^2/df_{\text{diff.}}$ ($p < 0.01$) and the following fit indices: RMSEA: 0.061, SRMR: 0.0748, IFI: 0.918, TLI: 0.910, and CFI: 0.918.

8. We computed the main variables and their pair-wise interactions and correlated them in SPSS 23.

9. Following Tanriverdi and Venkatraman’s (2005) procedure, we did not include the performance variable in this test.

10. As suggested by Camacho-Minano et al. (2013), we controlled for size and unionization. Size was proxied for by the number of facility employees and facility sales, and respondents were asked to indicate whether their facility was fully unionized or not. We ran tests with respect to both hypotheses where size variables were additively related to firm performance and $\chi^2$ difference tests where size variables moderated all structural relationships. We ran the same tests regarding unionization. We find that all statistical inferences remain similar across all tests.

References


Crocker, L. and Algina, J. (1986), Introduction to Classical and Modern Test Theory, Harcourt, Brace and Jovanovich, Forth Worth, TX.


Cunningham, J.E. and Fiume, O.J. (2003), Real Numbers: Management Accounting in a Lean Organization, Managing Time Press, Durham, NC.


Emiliani, M.L., Stec, D. and Grasso, L. (2003), Better Thinking, Better Results, The Center for Lean Business Management, LLC, Kensington, CT.


Appendix

<table>
<thead>
<tr>
<th>Measure</th>
<th>Test variable</th>
<th>Explanation for correlation</th>
<th>Properties test variable</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>OUTNF</td>
<td>Cost of quality</td>
<td>If a firm uses non-financial control mechanisms, it is likely to measure the cost of quality</td>
<td>Single item</td>
<td>0.583**</td>
</tr>
<tr>
<td>OUTNF</td>
<td>Productivity</td>
<td>If you use non-financial management control mechanisms, you are likely to measure productivity</td>
<td>Single item</td>
<td>0.515**</td>
</tr>
<tr>
<td>OUTNF</td>
<td>On-time deliveries</td>
<td>If a firm uses non-financial management control mechanisms, we expect it to measure on-time deliveries</td>
<td>Single item</td>
<td>0.431**</td>
</tr>
<tr>
<td>OUTNF</td>
<td>First-pass yields</td>
<td>We expect that if a firm uses non-financial control mechanisms, it is likely to measure first-pass yields</td>
<td>Single item</td>
<td>0.538**</td>
</tr>
<tr>
<td>OUTNF</td>
<td>Cycle time improvements</td>
<td>If a firm uses non-financial management control mechanisms, it is likely to measure cycle time improvements</td>
<td>Single item</td>
<td>0.573**</td>
</tr>
</tbody>
</table>

Note: **Significant at $p < 0.01$ (two-tailed)

Table A1. Criterion validity for output non-financial control mechanisms

Complementary management control mechanisms
Social cultural controls
Please indicate below what most closely represents your facility’s organizational culture
1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree
CLTR 1 Management style is more participative than autocratic
CLTR 2 Management is committed to quality-related training
CLTR 3 All employees are involved in problem solving
CLTR 4 Our entire facility is trained in lean principles
CLTR 5 Every area of our facility works on continuous improvement
CLTR 6 Management is focused on eliminating waste everywhere
CLTR 7 Lean thinking has permeated all of our operations
CLTR 8 Team members feel peer pressure to perform

Social visual controls
For the following items, please mark the most appropriate response related to your facility’s management accounting system
1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree
MAS 1 Standard operating procedures are visible on the shop floor
MAS 2 Visual boards are used to share information
MAS 3 A training skills matrix is visible on the shop floor
MAS 4 Charts showing defect rates are posted on the shop floor
MAS 5 We have created a visual mode of organization
MAS 6 Information on productivity is updated frequently on the shop floor
MAS 7 Quality data are displayed at work stations

Behavioral controls
Please indicate below the extent to which your facility has implemented the following
1: not at all, 2: little, 3: some, 4: considerable, 5: great deal
MFG 1 Use of standardization
MFG 2 A Kanban system
MFG 3 Use of one-piece flow
MFG 4 Use of line balancing and level schedules

Non-financial output controls
Please indicate below how important these performance measures are to operations at your facility
1: not at all, 2: somewhat, 3: important, 4: very important, 5: critical
PRF1 Non-financial measures related to cell performance
PRF2 Non-financial measures related to value stream performance
PRF3 Non-financial measures related to facility performance

Financial output controls
Please indicate below how important these performance measures are to operations at your facility
1: not at all, 2: somewhat, 3: important, 4: very important, 5: critical
PRF 4 Market share
PRF 5 Cash flow
PRF 6 Overall financial results
PRF 7 Customer satisfaction

Firm performance
Please indicate to what extent lean initiatives have affected the following
1: not at all, 2: little, 3: some, 4: considerable, 5: great deal
LIMP 1 Inventory-related resources have been freed up
LIMP 2 Capacity is managed more effectively
LIMP 3 Cycle/production time is improved
LIMP 4 Quality is improved
LIMP 5 Overall communication is improved
LIMP 6 Costs are reduced
LIMP 7 Profitability is improved

Table AII.
Survey items

Corresponding author
Henrik Nielsen can be contacted at: henn@business.aau.dk

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Controlling lean manufacturing in multidivisional organisations
Highlighting local interests and constraints

Martijn Pieter van der Steen and Sandra Tillema
Department of Accounting, Rijksuniversiteit Groningen, Groningen, The Netherlands

Abstract
Purpose – The purpose of this paper is to address the impact of a multidivisional structure on the implementation of lean manufacturing. It investigates how the controls employed by the corporate level impact the local implementation of lean manufacturing.

Design/methodology/approach – The paper reports on case studies in three subsidiaries in different multidivisional organisations.

Findings – The paper finds that lean manufacturing can be severely constrained by the accounting-based controls which are commonly in place in a multidivisional structure. Depending on the degree of centralisation, subsidiaries may be restricted to implementing lean tools in a fragmented way, rather than acting according to a coherent set of principles.

Practical implications – Companies may have to accept that being part of a multidivisional organisation can imply that their lean implementation is more gradual and piecemeal than they prefer. The paper proposes several ways to mitigate the constraints that may arise from incompatibilities between accounting-based controls and lean controls.

Originality/value – This study contributes to the literature about external constraints on production innovations, such as lean manufacturing. It highlights how the organisational context creates local conditions that may be detrimental to the implementation of lean manufacturing.

Keywords Lean, Management control, Lean manufacturing, Control system, Multidivisional structure

1. Introduction
Since the 1990s, lean manufacturing has attracted the attention of western manufacturing companies as an approach to achieve superior performance (Hines et al., 2004; Holweg, 2007). However, despite its positive effects on manufacturing performance (Cua et al., 2001; McKone et al., 2001; Shah and Ward, 2003), lean manufacturing has always had an uneasy relation with the systems through which organisations have traditionally controlled their performance. These traditional systems rely heavily on accounting-based controls. Already in the 1980s, researchers observed that such control systems are unable to support modern manufacturing practices (Kaplan, 1984, 1989; Brimson and Berliner, 1987). They argued that these systems produce inaccurate cost price information, and ignore many of the benefits that arise from excellent performance in the areas of quality, flexibility, throughput time and customer responsiveness. Consequently, they continued, the traditional control systems make it difficult for organisations to understand which actions are needed to improve manufacturing performance.

In lean environments, control systems should reward managers and employees for efforts which reduce waste and increase quality (cf. Fullerton and McWatters, 2002). These systems may be inconsistent with the extant accounting-based controls, which are predominantly financial in nature (Hansen et al., 2003; Li et al., 2012). These financial
controls may promote actions which improve financial performance, but which, for instance, decrease quality or lead to increased inventories. Because of the distortions created by accounting-based controls, various authors suggest that companies must adjust their control systems in order to fully reap the benefits of lean manufacturing (e.g. Åhlström and Karlsson, 1996; Bhasin, 2008; Hope and Fraser, 2003; Li et al., 2012; Maskell et al., 2012). These adjustments can lead to control systems which motivate behaviour that is more aligned with the principles of lean manufacturing (Bhamu and Sangwan, 2014). Empirical findings confirm that such adjustments are essential. For instance, Fullerton and Wempe (2009) find that utilising non-financial performance measures is a requirement for lean to be successful in financial terms. Various studies also show that companies which have implemented lean practices have adjusted their control systems; for example, by including more non-financial and bottom-up measures of performance (Baines and Langfield-Smith, 2003; Fullerton and McWatters, 2002).

However, previous studies have paid limited attention to the implications that a multidivisional structure may have for these adjustments of control systems. Most studies either investigate independent companies (e.g. Chiarini, 2012; Fullerton and Wempe, 2009; Powell et al., 2013), or largely ignore the consequences of multidivisional structures for the control of manufacturing facilities (e.g. Kennedy and Widener, 2008; Fullerton and McWatters, 2002). This emphasis in previous studies may be a reflection of how companies regard lean implementations. Hines et al. (2004) find that companies pay much attention to localised shop-floor implementations of lean manufacturing, even to the extent that this has resulted in “island optimisations”, rather than a holistic organisation-wide perspective. Bhasin (2012) observes that mechanisms which hold organisational activities together, such as performance measurement and performance compensation systems, are often omitted from lean implementations. As a result, the initial gains of lean manufacturing often remain localised (Mohanty et al., 2007). Such a lack of immediate impact on performance measures that are important to higher-level managers can be problematic, because it could result in the loss of support from higher levels for a company’s lean implementation (Li et al., 2012).

Hence, the control of the production system does not take place in isolation. Rather, it is embedded in wider organisational controls, which are often based on accounting measures (Kennedy and Widener, 2008). This especially holds for multidivisional organisations. In these organisations, the corporate level controls the activities of the lower levels predominantly through accounting-based controls (Hansen et al., 2003; Hope and Fraser, 2003; Østergren and Stensaker, 2011). Examples of such controls are budgets and financial performance indicators (e.g. net income, return on investment). The effects of the corporate level’s use of such controls on innovative manufacturing technologies have not received much attention in previous studies.

This study investigates how lean manufacturing in multidivisional organisations is affected by the extant infrastructure of accounting-based controls. It is based on case research in three production companies in the Netherlands. These companies implemented lean manufacturing, which significantly altered their production function. In these companies, several frictions were identified, which arose from incompatibilities between the controls employed at their corporate levels and the needs of lean manufacturing. Through the analysis of these frictions, the study demonstrates that demands from the corporate level can constrain subsidiaries in their attempts to follow the principles of lean. The study also proposes ways to mitigate these constraints.

The remainder of this paper is structured as follows. Section 2 discusses management control in a multidivisional setting, and the incompatibilities between the accounting-based and lean controls. Next, Section 3 describes the research method employed for this paper. Subsequently, Section 4 presents the findings from the case studies, and Section 5 discusses their implications. Finally, Section 6 presents the conclusions as well as some directions for future research.
Lean manufacturing and management control in multidivisional organisations

Although lean manufacturing systems can bring many benefits, there are various hurdles which may threaten the implementation and continuance of lean in organisations (Turesky and Connell, 2010; Taylor et al., 2013). To overcome such hurdles, previous studies have generated a range of factors that have been shown to be critical to the success of lean manufacturing systems (e.g. see Achanga et al., 2006; Bhasin and Burcher, 2006; Worley and Doolen, 2006). These studies may give rise to the idea that organisations are able to avoid lean failures by carefully attending to these success factors. For instance, Turesky and Connell (2010, p. 111) observe that:

"Given that the approach to lean manufacturing initiatives is well-documented, it seems odd that consultants and their counterparts in leadership would fall victim to such threats to sustainability [of lean manufacturing]."

However, lean implementations may also be subject to extraneous constraints, which cannot be directly influenced. Bamford et al. (2015) argue that partial implementation of lean is not always the result of deliberate organisational choice or a lack of a firm belief in lean. Instead, extraneous constraints may force an organisation to implement lean in a "more patchy, piecemeal and partial" way (p. 703).

Previous studies have considered constraints related to, for instance, seasonal patterns in the demand for products and supplier reliability (Bamford et al., 2015) and dissimilar socio-economic and cultural conditions between countries (see Moyano-Fuentes and Sacristán-Díaz, 2012). This paper aims to add to this body of research by arguing that – from a subsidiary perspective – the multidivisional structure can also be a source of constraints which can be influenced only to a limited extent (see e.g. Maalouf and Gammelgaard, 2016).

The benefits of a multidivisional structure have been described at length (see Chandler, 1962), and these descriptions mostly highlight its ability to minimise transaction costs and thus to achieve an optimisation of organisational efficiency (Williamson, 1970, 1975, 1981). Accordingly, the so-called M-form hypothesis predicts that large multidivisional organisations produce superior outcomes in comparison to other organisational forms. These outcomes are mainly attributed to the high level of decentralisation and autonomy of decision making at the local level.

Yet, there is conflicting evidence about the benefits of decentralisation of decision rights and plant autonomy for implementing lean manufacturing. Lower degrees of decentralisation are found to be associated to better alignment of organisational goals, simpler decision making and better coordination (Puranam et al., 2006). These factors facilitate the roll out of lean manufacturing. By contrast, recent findings suggest that higher degrees of decentralisation facilitate a sense of ownership of the lean implementation and greater plant management autonomy in decision making about lean manufacturing (Secchi and Camuffo, 2016). Secchi and Camuffo (2016) tentatively conclude that the benefits of a higher degree of managerial autonomy and localised decision making more than offset the advantages of tighter control by the corporate level. Therefore, one may expect that multidivisional organisations provide an appropriate context for implementing lean manufacturing.

However, although multidivisional organisations are typically associated with decentralisation, some degree of centralisation is needed to hold the different parts of the organisation together (Otley and Berry, 1980). To enable the corporate level to control its subsidiaries, multidivisional organisations use elaborate systems of accounting-based controls. Flamholtz (1996) observes that these systems permit corporate managers to delegate day-to-day operations to lower-level managers, while simultaneously assuring that these managers remain focused on the strategic goals of the organisation (see also Lowe, 1971; Otley, 1999). In addition, authors such as Hansen et al. (2003) and Otley and Berry (1980) indicate that accounting systems enable the corporate level to integrate the management control system with the needs of stakeholders such as shareholders and banks.
Accounting-based controls introduce new forms of centralisation, as organisation-wide “centres of calculation” are used to control all subsidiaries on similar criteria (Quattrone and Hopper, 2005). Hence, through their use of accounting-based controls, multivisional organisations can establish varying levels of centralisation (Pellinen et al., 2016). However, the use of accounting-based controls may be detrimental to the success of lean implementations for two reasons. First, as previously argued, the extensive use of accounting-based controls may result in a de-facto centralisation of decision-making authority. And second, accounting systems were originally developed for mass production systems, which were slow to change and inflexible by nature. Hence, accounting systems may not be appropriate to respond to new environmental demands, including those emanating from new technologies (Pirson and Turnbull, 2015; Hamel and Prahalad, 1983). Therefore, the use of accounting-based controls in multivisional organisations may negate the aforementioned benefits of decentralisation.

Modern production technologies are not a mere set of tools and practices, such as total quality management, a Kanban system and cellular manufacturing (for an overview, see Shah and Ward, 2003). Instead, in order to be successful, these tools and practices should be tied together into a complete system (Womack and Jones, 1996). The backbone of such a system is a set of principles, which should guide all actions. In the context of lean, these principles include creating value from the perspective of the customer, and aiming at perfection (Womack and Jones, 1996). Following such principles is facilitated by a set of local controls, which include specific performance measures, standard operating procedures and other, more social, ways to control workers (Fullerton et al., 2013; Kennedy and Widener, 2008; Kristensen and Israelsen, 2014). However, various authors have argued that the actions promoted by these local controls may not be consistent with accounting-based controls (e.g. Kaplan, 1984, 1989; Brimson and Berliner, 1987; Maskell et al., 2012).

Hansen and Mouritsen (2007) draw on these authors to provide a systematic classification of these inconsistencies. They identify four incompatibilities between accounting-based controls and controls which support modern production technologies, including lean manufacturing. Table I uses this classification to structure the literature about the implications of using accounting-based controls in lean environments.

Table I shows that a misalignment between accounting-based and lean controls impedes the successful implementation of lean (see also Åhlström and Karlsson, 1999). As mentioned earlier, organisations often adjust their control systems to deal with the incompatibilities between accounting-based and lean controls. The control systems in lean companies usually rely more heavily on non-financial and bottom-up measures of performance (Baines and Langfield-Smith, 2003; Fullerton and McWatters, 2002), and they place more emphasis on behavioural and social controls, such as standard operating procedures, empowerment and training (Kennedy and Widener, 2009). In addition, the accounting systems in these companies are more simplified and streamlined, and they tend to focus on value streams (Fullerton et al., 2013).

However, in a multidivisional organisation, subsidiaries are not commonly able to influence the control system used by the corporate level. Therefore, implementing lean manufacturing in a multidivisional organisation presents a specific set of challenges, which are related to the degree of centralisation and the use of accounting-based controls by the corporate level. The ways in which these challenges unfold are expected to affect the success of lean implementations. Hence, this paper problematizes the implementation of lean in multidivisional organisations. It confronts the merits of an autonomous local implementation and operational control with the varying degrees of centralisation brought about by the use of accounting-based controls. To this end, the following two research questions (RQs) are formulated:

**RQ1.** How does the corporate level’s use of accounting-based controls in multidivisional organisations constrain the local implementation of lean manufacturing in their subsidiaries?
RQ2. How can variations in the level of centralisation explain differences in the success of lean implementations between subsidiaries of different multidivisional organisations?

The paper distinguishes between the four areas of incompatibility identified by Hansen and Mouritsen (2007), because these represent areas where local and corporate interests can collide. Hence, these areas are fundamental sources of constraints on lean manufacturing in multidivisional organisations. The aim of the study is to explore how companies, which are part of a multidivisional structure, can increase the success of their lean implementation.

3. Research method

An inductive case study (Barratt et al., 2011) was conducted. This research method is particularly suitable to understand context and practitioner experiences (Fisher, 2007).
In general, such a case study approach fits well with the type of RQs as formulated in Section 2, which require a thorough understanding of the nature and complexity of the phenomenon under study (Voss et al., 2002). The cases selected for this research were three Dutch production subsidiaries, which will be referred to as Midden, West and Zuid. Some of the features of these subsidiaries are listed in Table II.

These subsidiaries were selected because they were all part of multidivisional organisations which operated an elaborate accounting-based control system. The subsidiaries varied in the extent to which their corporate level, operationalised as headquarters (HQ), intervened in local strategy and operations. Hence, they varied in the level of centralisation. The case selection was thus based on a theoretical sampling strategy (Barratt et al., 2011).

The empirical research was conducted in three stages. The first stage included an interview with a lean consultant and a discussion meeting with representatives of Midden and West[1]. During the discussion meeting, the representatives of the two companies gave an understanding of the problems related to the control of their lean operations. In the second stage of the study, two research assistants were enlisted to spend three months at each of these companies to explore the relationships between the introduction of lean manufacturing and the extant management controls. For this purpose, they interviewed and had informal talks with persons at all levels in their company and they collected internal documents. They conducted a total of 26 interviews of varying length. The third stage took place during the same timeframe, in which the researchers interviewed a total number of 12 managers, consultants and management accountants in Midden, West and Zuid. Some persons were interviewed on more than one occasion. The average length of the interviews was 1:55 hours. All interviews are listed in Table A1.

The interviews dealt with the implementation of lean manufacturing, the design and use of the management control system (including the accounting system) in the “new” lean manufacturing environment, and the relation with HQ. Additional interviews were added until theoretical saturation was reached; that is, when newly analysed data did not prompt

<table>
<thead>
<tr>
<th>Table II. Features of case sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry</td>
</tr>
<tr>
<td>Industry</td>
</tr>
<tr>
<td>Total no. of employees</td>
</tr>
<tr>
<td>Global sales (Million EUR)</td>
</tr>
<tr>
<td>No. of production subsidiaries</td>
</tr>
<tr>
<td>No. of employees in subsidiary</td>
</tr>
<tr>
<td>Reason for implementing lean</td>
</tr>
<tr>
<td>Level at which initiative for lean was taken</td>
</tr>
<tr>
<td>Stage of lean implementation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Midden</th>
<th>West</th>
<th>Zuid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of centralisation</td>
<td>High</td>
<td>Medium</td>
</tr>
</tbody>
</table>
further changes to the theoretical understanding of the events in the companies (Karlsson et al., 2010). This approach ensured comparability of the data, because more data were added until the three case studies could be compared on the four themes classified in Table I, their histories of implementing lean and the ways they were controlled by their corporate levels. Following some of these interviews, persons at the three companies invited the researchers for a site visit. These visits were helpful for obtaining an understanding of the operational implications of lean. The data collection process ended with a meeting with representatives of the case companies to discuss the initial findings of the study. Together with the interviews and the observations during the site visits, the discussion meetings also offered opportunities for triangulation of the data and corroboration of the conclusions (Barratt et al., 2011; Voss et al., 2002).

The data analysis followed several of the suggestions made by Voss et al. (2002). All interview recordings were coded and then analysed according to the classification by Hansen and Mouritsen (2007). Each theme in this classification emphasises a specific set of constraints that the subsidiaries encountered due to the controls that were imposed by HQ. In order to understand the origins and consequences of these constraints as well as the conditions under which they could be mitigated, the researchers analysed their summaries and coded transcripts in detail and revisited the relevant parts of their recordings. Also the relationships between the three companies and their corporate levels were compared, to understand the context that had given rise to a particular level of centralisation.

4. Empirical findings
This section presents the findings from the case studies. It is organised around the two RQs.

4.1 Impact of HQ controls on lean implementation
One of the cornerstones of the multidivisional form is a set of accounting-based controls, which binds together the different units and levels of the organisation (Hansen et al., 2003; Hope and Fraser, 2003; Østergren and Stensaker, 2011). As lean is often initially limited to production units, it deploys controls which are restricted to the local level (Hines et al., 2004; Bhasin, 2012). This section will demonstrate how each of the four incompatibilities between both types of controls affected the implementation of lean.

Financial vs non-financial performance measures. In the case companies, performance measures did not cascade down the organisational hierarchy. The companies combined the use of lean-oriented performance measures at the lowest level with the use of a traditional budgeting system by HQ. This traditional budgeting system was largely financial in nature, and put much emphasis on standard cost variances; for instance, related to the utilisation of machines. It was considered a benefit that the financial measures brought in a shareholder perspective. For example, when asked which performance measures receive most attention, a controller in West said:

O yes, that’s an easy one! From headquarters? Conversion cost [variances], I believe. Yes, and that is because, in the end, also from their perspective, if they are evaluated, conversion costs are of primary importance. That is not to say that [other performance measures] are irrelevant, but I believe that in this respect our company is financially driven. Whether you like it or not.

The financial measures were not always consistent with the lean-oriented performance measures used by the subsidiaries, which were more operational in nature, and which led to a heightened focus on the customer. A former senior manager of Zuid suggested that frictions between different sets of performance measures are typical of multidivisional companies. He argued:

[In a multidivisional organisation], I am very restricted in the accounting measures I can use. The emphasis on contribution is at the expense of a focus on the customer. If you have a machine,
its costs need to be recovered; however, lean does not look at contribution, but at sales to the end customer. […] It leads to massive conflicts when you state that you are not really interested in contribution.

When attempting to implement lean in a [multidivisional] organisation, many aspects are not really open for discussion. One of these is attempts to improve [the format of] monthly reports. But if these reports are not changed, you will not have discussions about customer value and the benefits of investments in lean, so you cannot show important things.

In Midden and West, lean was successful in operational terms, but not in financial terms. In these companies accounting had lost its role as an “integrative device” into which all the activities of the organisation can be “drawn together” (cf. Otley and Berry, 1980, p. 234). One explanation for the lack of a positive impact of lean on financial performance in Midden and West was that improvements in financial performance were postponed by re-investing the savings that resulted from lean. However, this strategy was considered risky, because it could lower HQ’s support for lean.

Particularly in Midden, inconsistencies between the two types of performance measures constrained the implementation of lean (cf. Bamber and Dale, 2000). The company had an increasing number of conflicts with the managers at HQ due to the different understandings about performance and how to measure it. There were various instances where the traditional variance metrics were adversely affected by decisions which were beneficial from a lean perspective. For example, the demand for the company’s products decreased sharply as a result of the recent recession in Europe. This reduction in demand prompted a manager at the subsidiary level to call for a reduction of the output of the subsidiary. However, such a reduction in output would lead to a sharp increase in unfavourable variances. The lower production volume would immediately be translated into losses, as the fixed costs of the expensive capital goods could not be fully absorbed by the units actually produced. This was not acceptable for the managers at HQ, who did not see the need for lower production volumes. They asked the subsidiary to maintain production levels, and to store the surplus output temporarily in their warehouse. Therefore, there was an incentive to maintain or increase production, even when there was no demand. A respondent explained:

By increasing production we can easily perform better than the budget. The result is that you end up with stock, but it looks very favorable on the cost calculations. When there is pressure for better results it is very tempting to say: we will increase production.

Hence, HQ’s requirement to meet financial targets forced Midden to give up some of the lean principles.

**Status quo vs improvement.** Lean manufacturing promoted innovation in the three companies. Lean initiatives fundamentally changed the nature of production processes, which led to lower lead times, less waste and higher quality. However, the decision-making tools favoured by HQ induced a more conservative behaviour. By focussing on absorption costs, relevant costs and revenues, and incremental cash flows, these tools were often in favour of investments in existing production technologies. A former senior manager in Midden noted:

It is not in the nature of lean to account for its benefits in the way that HQ likes to see. The whole point is to cease the wasteful collection of data. We know there are many benefits to working smarter; this is purely logical. However, to put a number on those is difficult. Moreover, the gains we obtain from lean operations will be used for further improvements in our processes, there may therefore not be a direct financial gain, but certainly many benefits in the quality of our processes and products.

Following the lean principles, many interviewees took the view that it was important for their company to reduce lead and change-over times, to improve product and process quality, and to empower employees. They were convinced that such initiatives would have
positive financial consequences, but neither before nor after their implementation, they were able to quantify these consequences. Zuid did not consider major investments at the time, but in Midden and West, people expressed their frustration that they could not convince HQ of the benefits of investing in lean. According to an industrial cost engineer in West:

An issue which does generate discussion is, for example, this project I am working on (using a high labour intensity to increase flexibility). At some point we had to invest a sum of money. […] What you see then is that it is obviously a problem. "Come on, this is not China", and so on. Their viewpoint regarding this issue is perhaps a bit old school, so to say. You try to explain to them the importance of following the market, being flexible, being able to changeover quickly etc. However, the benefits of such an approach don't mean much compared to other things, it seems. Perhaps that is the problem; that you have to convince people. You may wish you already passed that stage. It is simply a matter of difference in mind-set.

Both Midden and Zuid started implementing lean manufacturing without informing their HQ. In this way, they attempted to avoid discussions about their inability to quantify the benefits of lean manufacturing. In Midden, lean initiatives were initially treated as a series of small changes in the production process. A former manager explained:

In the first year, HQ did not know that we were implementing lean. We undertook everything in small steps so there were not many costs involved. The costs that we had, we reported them as quality costs to our parent.

However, as the lean implementation grew and started to encompass both of the production lines, the management team of the subsidiary informed HQ about their efforts. The managers at HQ were mostly indifferent to these efforts. They were very keen on the efficiency gains that lean would yield, but they were opposing any up-front investment in the process improvement technologies of lean, as there was no business case which financially justified lean manufacturing. As a result of the limited resources available, the subsidiary continued with a low-key rollout of lean. Hence, the subsidiary was restricted to smaller-scale lean initiatives, which did not require HQ approval. Consistent with Boyer (1996), the low level of support for lean by HQ put a constraint on the resources available to invest in lean initiatives.

Hierarchical vs lateral relationships. At all companies under investigation, HQ exercised control over the lower levels largely based on accounting information about individual organisational units. These hierarchical controls enabled HQ to act at a distance (cf. Quattrone and Hopper, 2005). However, based on the lean principles, the subsidiaries had introduced lateral controls, i.e. controls based on value streams rather than organisational units. These lateral controls could not easily be reconciled with the existing hierarchical controls. In West, for instance, interviewees explained that when they invested in a reduction of lead times or change-over times, this would increase the investment base of the organisational units. This would have a negative impact on the return on investment, for which the managers of these units were held accountable.

In Midden in particular, the combination of different control orientations proved to be problematic. During a major reorganisation in 2006, HQ had imposed a matrix structure on this subsidiary. This structure induced departments to work independently. As a team leader illustrated:

Engineering does a lot by themselves and excludes other departments in their operations.

This matrix structure also meant that various departments, including the logistics and the procurement departments, were managed from abroad. According to a former manager who was interviewed, this matrix structure led to isolated “vertical pillars” running through the organisation, and the emergence of “islands” which strongly reduced cohesion in the facility.
This invoked discussions about the rationality of particular decisions. Decisions which were optimal from the perspective of a functional area destroyed value from the perspective of a value stream. For example, the former manager explained that they received large shipments of parts which were cheaper to procure in bulk from China. The manager had not enough space to store these parts locally and he lost the flexibility brought about by having low inventories. As a result, he was placed in a position which violated various of the principles of lean.

The geographical distance between the managers at HQ and the production operations was regarded as a real problem in the context of the lean philosophy employed in Midden. HQ attempted to benefit from their overview of similar operations across different subsidiaries, but they were hindered by a lack of knowledge about interdependencies within a single unit. On those occasions, Midden had to accept that it was constrained in its attempts to eliminate waste from its value streams (cf. Hansen et al., 2003). The subsidiary’s use of controls was rather unbalanced, as there was always an implicit choice for either lateral or hierarchical controls.

Top-down control vs empowerment. The events in the case companies expose a fundamental contradiction between trust in company-wide systems and trust in localised solutions. Whereas the traditional control systems operated by their HQs were mostly top-down and supposed to traverse geographical distances, the lean philosophy of the case companies dictated that control was localised and contextualised. To facilitate this localised type of control, the case companies had hung up screens in the production halls which displayed scores on key performance measures, and they had implemented a Kanban system which created visual controls on the work flows. Control was exercised by local managers, in particular those in charge of production. Moreover, since the introduction of lean manufacturing, notions such as shared goals and trust in the expertise and motivation of team leaders and workers, all had become more important in the case companies. A senior manager of Zuid observed:

It’s all very well, building a [company-wide] system and putting someone – who believes that he has the required knowledge too – in charge a mile away to operate it. But it is the people on the floor who really know what’s going on. They recognize the bottlenecks. If these people can be motivated, then you’re making progress. If you believe you must build a sophisticated control instrument around them, to force them with a gun, it will take you nowhere in the end.

The case companies believed that accounting information was useless to control their operations, and more and more they considered the activities needed to produce this information as waste (cf. Wallander, 1999; Maskell et al., 2012). However, HQ in all three companies continued to rely on the accounting system for controlling the lower levels. To keep this system up-to-date, Midden kept tracking its inventories of raw materials, work in progress and finished goods. By contrast, in West and Zuid, a Kanban system had replaced their planning for, and recording of, the production of individual components in the company-wide system. However, similar to the accounting function in the company studied by Ahlström and Karlsson (1996), management accountants and auditors in both West and Zuid resisted these attempts to eliminate waste. They were responsible for safeguarding the quality of the data needed to prepare reliable reports for HQ. In their view, this required a detailed tracking of internal transactions. According to a manager in West:

So regarding [the elimination of waste from our company-wide information system], every time we try to launch an idea [by saying:] "this is what we want to do", well, [the management accountants] start calling the head office claiming that the new manager has lost her mind, that she is doing crazy things. So it is a very sensitive issue, because knowledge gives people power. They think: as long as I know what this item costs, I can …. But it is a false sense of security that we have built.
West continued to use weekly inventory counts for its work-in-progress inventories to prepare the periodic financial statements required by HQ. By contrast, Zuid had decided to replace the regular inventory counts with a new type of internal controls. These controls relied on an orderly flow of products through the factory combined with visual controls on this flow. The internal auditors, however, did not accept this type of controls. According to a senior financial manager:

As long as you are able to print out lists from your [company-wide information] system, it is all right [according to auditors]. So auditors are simply confident that existing systems function properly. It is my experience that, although the system may function properly, its information is not really useful. However, it provides the auditor with a more solid basis, for it is a system which has been designed by an IT-consultant, who also arranged for its certification. So anything the system spits out is regarded as reliable, no matter what inputs were used. [Our approach] is based on the work of man, which means that mistakes can be made. But of course also the system depends on the input of people!

Hence, both West and Zuid were hindered in their attempts to eliminate “wasteful” activities, as these attempts were experienced as a threat to HQ’s ability to control the company.

To answer RQ1, Table III gives an overview of the frictions between accounting-based and lean controls experienced by the case companies at the time of the research. The table shows that, in these companies, HQ put much emphasis on financial performance. For this reason, measures of both financial performance and internal controls which had to ensure the reliability of financial information were key elements of the system through which HQ controlled the subsidiaries. In addition, Table III reveals that HQ’s reliance on these types of controls constrained the subsidiaries in their attempts to follow the lean principles. On various occasions, these subsidiaries were forced to apply “old-fashioned

<table>
<thead>
<tr>
<th>Incompatibility</th>
<th>Company</th>
<th>Origin</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial vs</td>
<td>Midden</td>
<td>HQ focuses almost exclusively on financial performance; adverse standard cost variances are not accepted</td>
<td>Subsidiary is prevented from fully applying operational controls according to lean principles</td>
</tr>
<tr>
<td>non-financial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Status quo vs</td>
<td>Midden</td>
<td>HQ only gives approval for investments if improvements in financial performance can be expected. Subsidiary is not able to calculate the effects of lean initiatives on financial performance</td>
<td>Up front investments in lean initiatives do not get approval from HQ. Subsidiary can only take small-scale lean initiatives</td>
</tr>
<tr>
<td>improvement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td></td>
<td>When assessing business cases, HQ puts much emphasis on improvements in financial performance. Subsidiary is not able to capture benefits from lean initiatives in traditional decision-making tools applied by HQ</td>
<td>Subsidiary finds it very difficult to convince HQ of the advantages of lean-oriented investments as compared to mass-production alternatives</td>
</tr>
<tr>
<td>Hierarchical vs</td>
<td>Midden</td>
<td>To maximise financial performance across different subsidiaries, HQ intervenes in subsidiary’s operations</td>
<td>Subsidiary is forced to ignore interdependencies within value streams; subsidiary is not fully able to focus on optimising value streams</td>
</tr>
<tr>
<td>lateral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top-down vs</td>
<td>West</td>
<td>HQ requires reliable information at any point in time</td>
<td>Subsidiary’s attempts to eliminate “wasteful” accounting tasks were opposed by HQ</td>
</tr>
<tr>
<td>empowerment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zuid</td>
<td></td>
<td>HQ requires the deployment of traditional internal controls</td>
<td>Subsidiary’s attempts to eliminate “wasteful” accounting tasks were opposed by HQ</td>
</tr>
</tbody>
</table>

Table III. Summary of origins and consequences of HQ’s use of controls
‘efficiency’ thinking” (cf. Womack and Jones, 1996, p. 18), while they would have preferred to create value for the customer, by offering more product variations and removing waste from operating and administrative processes. The controls applied by HQ resulted in larger batches and larger inventories, lower flexibility, and more resources devoted to keeping the company-wide information system up-to-date, than the subsidiaries deemed necessary and desirable. Table III also shows that there is a variation between the case companies in terms of the types of constraints experienced. Section 4.2 will explore how this variation is related to the degree of centralisation in each of the companies.

4.2 Variations in the level of centralisation and lean implementations
Of the three companies, Zuid was most unaffected by the incompatibilities listed in Table III. Although HQ was not particularly interested in lean, Zuid had been able to carve out local spaces for implementing lean. It had implemented lean without the permission of its HQ, but this implementation had positive effects on its financial performance. Especially the increased flexibility of manufacturing enabled it to compete with larger plants in Asia and sell its products at a premium. As a result, the company was able to meet the financial targets set by its HQ. A consequence was that HQ had no incentive to invoke additional accounting-based controls. Hence, Zuid operated as a highly decentralised unit, which had much decision autonomy on operational and strategic matters. This autonomy had enabled Zuid to expand and refine its lean operations to the extent that the plant not only operated a wide variety of lean tools (such as a Kanban system, regular 5S activities and value stream mapping), but also that it acted in line with the underlying principles of lean (e.g. by interrupting production when problems were detected).

In West, HQ supported the implementation of lean. Since HQ had decided to embrace lean manufacturing, the subsidiary experienced that there was no need to continuously renegotiate the relative importance of traditional performance measures and lean-oriented measures. HQ’s support for lean had translated into a willingness to relax some financial controls and consent to the use of some non-financial measures of performance. Respondents also indicated that in recent years it had become easier to focus on value streams. In this company, lean was not limited to a set of tools. Managers increasingly asked each other which actions were most appropriate in light of the lean principles. They acknowledged that HQ’s support for lean enabled them to bring up lean principles as a valid set of concerns at all levels of the organisation. However, it must be noted that there were occasional conflicts with higher levels, because West could not quantify the benefits of lean. Especially the central and local accountants could not be convinced that the accounting-based controls needed to be relaxed, if only temporarily.

Table III shows that Midden experienced the most frictions due to incompatibilities between lean and accounting-based controls. HQ of Midden had no interest in lean and it had imposed a large set of accounting-based controls, which reduced much of the autonomy of the subsidiary. Hence, despite being part of a multidivisional organisation, Midden was controlled in a highly centralised fashion and, as a result, it had limited decision authority. Midden was restricted to using isolated lean tools, rather than an integrated approach. For example, Midden mapped value streams, but it could not initiate process improvements based on those maps. Such improvements had short-term cost implications and were thus rejected by HQ. In general, the implementation of lean tools in Midden had not resulted in an adoption of the lean principles. Instead, it had used its limited decision space to implement lean in a fragmented fashion.

These exploratory findings provide insight into RQ2. All case companies had been able to implement various lean tools, such as value streams and particular 5S techniques. However, the principles of lean (such as a focus on perfection and the elimination of waste) were more difficult to implement and maintain in companies that were more strictly
controlled through accounting-based controls. Hence, the difficulties associated to lean implementations were not related to the use of lean tools, but rather to the incompatibilities between the principles of lean and centralised accounting-based controls. In this vein, one may tentatively conclude that accounting-based controls can displace the introduction of lean principles, leaving only a set of relatively incoherent lean tools. In general, multidivisional organisations may not offer the conditions which are beneficial for lean implementations, if they impose highly restrictive central controls.

5. Discussion
The findings of this study can be used to explore how companies can increase the success of their lean implementation, by diminishing the constraints which arise from the corporate level’s use of controls. This section presents the outcomes of those explorations. It distinguishes between suggestions for lean implementation strategies aimed at reducing the degree of centralisation, and a suggestion for an alternative organisational form which avoids the degree of centralisation usually associated with the multidivisional form.

5.1 Lean implementation strategies
The variations observed among the three case companies suggest ways to reduce the degree of de-facto centralisation, and thus mitigate the constraints which result from the accounting-based controls applied by HQ.

Smooth the impact on financial performance. The lack of particular constraints in Zuid suggests that subsidiaries can reduce the number of HQ interventions in local decisions by producing financial results which are satisfactory to HQ. In order to achieve this, a subsidiary may need to compromise on the degree of reinvestment of savings from lean. In this way, the benefits of lean in financial terms should become visible to HQ in an earlier stage. In addition, the subsidiary could stabilise production volumes in periods of temporarily low demand by producing units of a regularly demanded product for stock. This will avoid excessive drops in financial performance. Although these strategies are inconsistent with the principles of lean, the findings in Zuid, which applied both strategies, suggest that these strategies can help the company over time to implement lean further. Consequently, this study provides further support for Bamford et al.’s (2015) argument that a stepwise implementation of lean cannot always be avoided, and may bring benefits to the organisation similar to a full-scale implementation. As long as HQ is exercising strict control based on accounting information, subsidiaries will be constrained to lean initiatives which do not require the approval from HQ, as was demonstrated by the findings in Midden and – in an earlier stage – Zuid. Such smaller-scale initiatives can produce the first pieces of evidence of the success of lean, particularly when some financial savings can be reported. In this way, smaller initiatives can pave the way for a larger scale rollout. This argument is supported by the findings of Ahlström and Karlsson (1996), who showed that lean initiatives have to produce positive results in financial terms before changes in the control system are conceivable. Ultimately, the initiatives may lead to a manufacturing system which is more flexible and involves lower fixed costs. With such a system, financial performance is less sensitive to fluctuations in production volumes. Consequently, using accounting-based controls is less problematic.

Gradually adjust the control system. Many papers claim that top management support is important, because top management provides the financial resources needed for implementing at the lower level (e.g. Worley and Doolen, 2006; Moyano-Fuentes, Sacristán-Díaz, 2012; Scherrer-Rathje et al., 2009). However, in the case of West, HQ support brought about a lower emphasis on accounting-based controls, which represented a de-facto decentralisation of decision autonomy (cf. Boyer, 1996). Decentralisation offers leeway for a more beneficial mix of controls for lean. To deal with the limited measurability of the effects of lean initiatives, the
assessment of investment proposals could gradually shift from a focus on the outcomes of traditional decision-making tools, to an approach which increasingly considers the expected, but difficult to quantify, improvements in operating processes. In the post-implementation phase, the effects of the initiatives on the subsidiary’s performance could be assessed using a combination of financial and operational performance measures, where the emphasis on operational performance may grow over time. Both strategies were applied in West, although the lean proponents in this company felt that the emphasis of HQ was still too much on traditional tools and measures. The findings in West also suggest that multidivisional organisations can learn how to deal with the incompatibilities identified between accounting-based and lean controls (e.g. Maskell et al., 2012; Chiarini, 2012; Hansen et al., 2003), provided that the corporate level and the subsidiary recognise the value of each other’s standpoint.

Combine different control orientations. If both satisfactory financial results and HQ support for lean are lacking, hierarchical forms of control are likely to be dominant. In the case of Midden, HQ exerted hierarchical power on many lower-level decisions, which hindered the optimisation of local value streams. Subsidiaries may then strive for a more balanced application of hierarchical and lateral controls to mitigate the constraints on implementing lean initiatives. Such an application of controls would imply that HQ keeps using its accounting-based controls, but acknowledges that financial results can sometimes be improved by considering lateral relationships. This can be achieved in two ways. First, HQ could involve subsidiaries in its decisions regarding interventions in particular functional areas. The subsidiaries can provide HQ with information about the consequences of particular interventions for their value streams, which should lead to more balanced decisions. Second, HQ could stimulate the expansion of lateral controls beyond its subsidiaries. For example, benchmarking could be used to share information about best practices and stimulate cooperation in improvement trajectories. Hence, where Kristensen and Israelsen (2014) showed that a balanced use of lean controls has a complementary effect on performance, the outcomes of this study suggest that there should also be a balance between these lean controls and the accounting-based controls that are important to HQ.

Find alternative ways to safeguard the reliability of accounting information. Irrespective of the subsidiary’s financial results and HQ’s degree of support for lean, keeping the company-wide information system up-to-date is likely to remain an important issue in multidivisional organisations. In all companies under study, HQ prevented a full erosion of its ability to control its subsidiaries. Hence, although there was considerable variation in the degree of centralised control, each multidivisional organisation under study retained a basic level of centralised control. In this vein, West’s and Zuid’s efforts to eliminate “wasteful” accounting activities created control problems for their HQ. To overcome such a deadlock, the accounting function and the operations domain together could search for combinations of internal controls which safeguard the quality of information without overly constraining the efforts to reduce “waste”. The outcome is likely to be a combination of accounting-based controls and controls built into operating processes. At the time of the study, Zuid followed this route. However, the company experienced that it was not easy to convince persons outside the operations domain of the effectiveness of operational controls. Researchers, such as Kennedy and Widener (2008) and Fullerton et al. (2013), observed that the implementation of lean manufacturing is associated with an elimination of activities needed to track internal transactions. The findings of this study reveal a need to examine more critically the impact of eliminating such activities on higher management levels’ abilities to control the lower levels. Developing internal controls which fit the lean principles seems feasible (see Maskell et al., 2012), but convincing accountants and auditors of their reliability might be challenging. Accountants and auditors have to become familiar with ideas about control
which exist outside their field of expertise, and they may be afraid of losing control (Åhlström and Karlsson, 1996). More generally, the challenges faced by the case companies link back to Chenhall’s (2003) call for research into the ways in which large, decentralised organisations combine controls to achieve sufficient degrees of diversification (i.e. fitting the local context) and integration (i.e. meeting central level demands).

5.2 An alternative organisational form

The mismatch between lean and the multidivisional structure may point to a need to consider alternative organisational forms which rely on controls that have a better fit with the operational emphasis of lean. We tentatively suggest that the so-called N-form, or network form, provides this better fit. Although often associated with collaborations between companies (e.g. Pekkola and Ukko, 2016), the network form also refers to specific ways of organising in a single organisation. It emphasises the combining of knowledge rather than the division of technologies and has the following characteristics (Hedlund, 1994). It entails temporary constellations of people, which provide operational flexibility. The network form highlights the importance of personnel at lower, operational levels and it promotes lateral communication. It promotes a catalytic and architectural role for top management, which is enabling rather than restrictive, and it emphasises focus, rather than economies of scale, as a competitive strategy. Finally, the network form recognises that a traditional hierarchy may be too restrictive, and thus different ways of ordering organisational activities may be deployed. Networked organisations combine various non-accounting-based controls which hold the organisation together, with a high level of local decision autonomy (Ruggero et al., 2016).

Secchi and Camuffo (2016) suggested that the implementation of lean manufacturing systems is positively associated with plant autonomy and that such implementation involves processes of experimentation and trial and error. They also highlight how knowledge sharing facilitates the effective implementation of lean. The findings in this paper extend these results by illustrating how the multidivisional form may not create the organisational conditions for the effective sharing of knowledge and local experimentation. The analysis demonstrates that the accounting-based controls in multidivisional organisations can be applied in a highly restrictive manner, discouraging the behaviours which are positively associated to lean implementation success. By contrast, the network form of organisations creates conditions that facilitate knowledge sharing and local autonomy. It supports information sharing as multiple relatively independent production units seek frequent collaboration with their peers to obtain best practices across the organisation. In addition, the network form can be controlled through mostly non-financial efficiency and quality measures and improvement actions can be initiated through instruments, such as benchmarking and value streams. The adoption of a network form may be only one tentative suggestion to avoid the detrimental effects of a strict use of accounting-based controls. In a more general sense, the findings of this paper suggest that researchers and managers may attend more to the ways the corporate level affects the local conditions, which, in turn, facilitate or impede the implementation of lean manufacturing systems. In this study, the focus was on accounting-based controls as influence on these local conditions, but it is likely that various other variables also impinge on these conditions.

6. Conclusion

Despite various innovations, both in accounting and in production, traditional accounting-based control systems are still in use in many organisations. In the multidivisional organisations included in this study, HQ persisted in using traditional accounting information to control the subsidiaries, partly because external stakeholders held them accountable based
on this information. Lean manufacturing provided a competing set of controls, which sometimes conflicted with the more traditional ones. The interferences of HQ based on information which was not consistent with lean manufacturing, constrained the subsidiaries in their efforts to follow the lean principles. In addition to constraints such as market characteristics and geographical context (Bamford et al., 2015; Moyano-Fuentes, Sacristán-Díaz, 2012), this study demonstrates that the multidivisional structure of many organisations may also introduce constraints on lean implementations; especially related to their use of centralised controls. This study provides further contributions by suggesting four different strategies to mitigate these constraints. Through these strategies, the case companies tried to adjust the system through which they were controlled, but this could only be done cautiously, to avoid any disturbances in their relationship with HQ. As such, the multidivisional setting adds extra complexities to lean manufacturing. A more radical, longer-term strategy may therefore be to introduce changes in the organisational structure.

The findings of this paper give rise to a more fundamental question related to the effect of extraneous conditions on the implementation of lean manufacturing. More specifically, there is a need for a richer understanding about the ways in which these conditions encourage or restrict the implementation of lean manufacturing. This paper highlighted the impact of the relation between different organisational levels. Yet, other conditions may be considered, including institutional factors, such as the requirements of providers of financing and external auditors, and labour laws which constrain the use of a flexible workforce. The variety of other constraints (see Bamford et al., 2015) emphasises the need for a coherent research programme that aims to understand and eliminate these constraints.

In this vein, this paper represents a call for further contributions to a broader understanding of the control of lean operations. We believe that it is important to appreciate the intricacies of control in lean companies, given the sometimes conflicting demands and frictions they experience. The effectiveness of lean manufacturing is rooted in a wide array of contextual issues, which this paper has only just begun to uncover.

Note

1. During this interview, it was pointed out that Zuid had implemented lean in a highly effective manner. Therefore, it was added to the study in a subsequent stage.

References


### Appendix

<table>
<thead>
<tr>
<th>Function</th>
<th>Company</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lean consultant</td>
<td>Ynova (consulting firm)</td>
<td>2:00</td>
</tr>
<tr>
<td>Lean consultant, industrial cost engineer West, industrial cost engineer no. 2</td>
<td>Ynova, West, Midden</td>
<td>2:00</td>
</tr>
<tr>
<td>West, lean consultant/former interim plant manager Midden, senior financial manager Midden, assistant financial manager Midden</td>
<td>Midden</td>
<td>2:30</td>
</tr>
<tr>
<td>Lean consultant/former interim plant manager Midden</td>
<td>Midden</td>
<td>2:30</td>
</tr>
<tr>
<td>Industrial cost engineer, industrial cost engineer no. 2, manager</td>
<td>West</td>
<td>2:30</td>
</tr>
<tr>
<td>Senior financial manager</td>
<td>Midden</td>
<td>2:00</td>
</tr>
<tr>
<td>Assistant financial manager</td>
<td>Midden</td>
<td>1:30</td>
</tr>
<tr>
<td>Former senior manager</td>
<td>Zuid</td>
<td>1:52</td>
</tr>
<tr>
<td>Senior manager</td>
<td>Zuid</td>
<td>2:39</td>
</tr>
<tr>
<td>Black belt lean manager</td>
<td>Zuid</td>
<td>1:47</td>
</tr>
<tr>
<td>Senior financial manager</td>
<td>Zuid</td>
<td>2:13</td>
</tr>
<tr>
<td>Senior manager, Black belt lean manager, senior financial manager</td>
<td>Zuid</td>
<td>0:45</td>
</tr>
<tr>
<td>Lean consultant/former interim plant manager Midden</td>
<td>Midden</td>
<td>1:15</td>
</tr>
<tr>
<td>Senior financial manager, assistant financial manager</td>
<td>Midden</td>
<td>1:15</td>
</tr>
<tr>
<td>Senior financial manager, assistant financial manager</td>
<td>Midden</td>
<td>1:50</td>
</tr>
<tr>
<td>Production manager no. 1</td>
<td>Midden</td>
<td>1:30</td>
</tr>
<tr>
<td>Production manager no. 2</td>
<td>Midden</td>
<td>1:00</td>
</tr>
<tr>
<td>Lean manager/Manager quality control</td>
<td>Midden</td>
<td>0:55</td>
</tr>
<tr>
<td>Supply chain manager</td>
<td>Midden</td>
<td>0:30</td>
</tr>
<tr>
<td>Technical project manager</td>
<td>Midden</td>
<td>0:17</td>
</tr>
<tr>
<td>Divisional procurement manager</td>
<td>Midden</td>
<td>0:35</td>
</tr>
<tr>
<td>Plant manager</td>
<td>Midden</td>
<td>1:45</td>
</tr>
<tr>
<td>Senior financial manager, assistant financial manager</td>
<td>Midden</td>
<td>1:47</td>
</tr>
<tr>
<td>Team leader production department</td>
<td>Midden</td>
<td>0:27</td>
</tr>
<tr>
<td>Manager product engineering</td>
<td>Midden</td>
<td>1:18</td>
</tr>
<tr>
<td>Lean consultant/former interim plant manager Midden</td>
<td>Midden</td>
<td>1:10</td>
</tr>
<tr>
<td>Assistant financial manager</td>
<td>Midden</td>
<td>0:42</td>
</tr>
<tr>
<td>Industrial cost engineer no. 3</td>
<td>West</td>
<td>0:30</td>
</tr>
<tr>
<td>Green belt programme manager</td>
<td>West</td>
<td>1:00</td>
</tr>
<tr>
<td>Maintenance employee</td>
<td>West</td>
<td>0:45</td>
</tr>
<tr>
<td>Employee technical support group line 2</td>
<td>West</td>
<td>0:45</td>
</tr>
<tr>
<td>Production manager hall 4</td>
<td>West</td>
<td>1:30</td>
</tr>
<tr>
<td>Lean coach</td>
<td>West</td>
<td>0:30</td>
</tr>
<tr>
<td>Quality inspector hall 2</td>
<td>West</td>
<td>0:45</td>
</tr>
<tr>
<td>Shift leader assembly line 2</td>
<td>West</td>
<td>0:30</td>
</tr>
<tr>
<td>Industrial cost engineer no. 1</td>
<td>West</td>
<td>1:40</td>
</tr>
<tr>
<td>Financial controller</td>
<td>West</td>
<td>1:31</td>
</tr>
<tr>
<td>Logistics capacity planner</td>
<td>West</td>
<td>1:00</td>
</tr>
<tr>
<td>Management team member supply</td>
<td>West</td>
<td>1:00</td>
</tr>
<tr>
<td>Employee logistics department</td>
<td>West</td>
<td>0:45</td>
</tr>
<tr>
<td>Meeting about the challenges of accounting and control in a lean environment</td>
<td>West, Midden, Zuid</td>
<td>4:00</td>
</tr>
</tbody>
</table>

*Note:* Indicates interviews conducted by the research assistants

---

**Corresponding author**
Sandra Tillema can be contacted at: s.tillema@rug.nl

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)
Or contact us for further details: permissions@emeraldinsight.com
The use of management control and performance measurement systems in SMEs
A levers of control perspective

Biljana Pešalj
Rotterdam Business School, Hogeschool Rotterdam, Rotterdam, The Netherlands
Andrey Pavlov
Cranfield School of Management, Cranfield University, Cranfield, UK, and
Pietro Micheli
Warwick Business School, University of Warwick, Coventry, UK

Abstract
Purpose – The purpose of this paper is to respond to recent calls for understanding how multiple management control (MC) and performance measurement (PM) systems are used simultaneously for managing performance, particularly in the context of small- and medium-sized enterprises (SMEs).
Design/methodology/approach – Data are collected during an in-depth case study of MC and PM and management practices in a Dutch SME using multiple data sources and elicitation methods, including interviews and participant observations.
Findings – This study identifies managerial practices that enable the interplay of the four control systems – beliefs, boundaries, diagnostic and interactive – helping the organization manage organizational tensions in relation to short- and long-term focus, predictable goal achievement and search for new opportunities, internal and external focus, and control and creativity.
Research limitations/implications – This paper advances the research on integrating multiple aspects of performance management, particularly technical and social. This research is based on a single case study; future qualitative and quantitative studies could explore the interplay between the four control systems in other settings and explore the relationship between control systems and leadership style.
Practical implications – Managing performance requires active and continuous use of all four control systems. This is particularly salient in SMEs where less formal controls play a key role and where balance needs to be ensured despite the lack of managerial processes and capabilities.
Originality/value – The findings advance PM and management theory and practice in the context of SMEs.

Keywords Performance measurement, SME, Performance management, Management control, Levers of control, Small- and medium-sized companies

Paper type Research paper

Introduction
Over the past two decades, the focus of research in management control (MC) and performance measurement (PM) has shifted from exploring the design and implementation of MC and PM systems to understanding how these systems are used and what their effects are (Hudson et al., 2001; Franco and Bourne, 2003; Pavlov and Bourne, 2011; Bititci et al., 2012). Various empirical studies have demonstrated that these systems can help organizations implement strategy, promote positive behaviors and improve performance both in the short and in the long term (Franco-Santos et al., 2012; Koufteros et al., 2014; Micheli and Mura, 2017). On the other hand, when consideration is mainly given to technical issues, such as the design of targets and measures, and social aspects are neglected (e.g. individuals’ understanding of the roles of measures and uses of performance information), effects of MC and PM systems tend to be negative (Smith and Bititci, 2017).

Importantly, the majority of empirical studies have been conducted in large companies and evidence is lacking in relation to how MC and PM systems are used in small- and
medium-sized enterprises (SMEs) (Bourne et al., 2000; Garengo et al., 2005; Brem et al., 2008; Bititci et al., 2012). This is problematic, not only because SMEs make a considerable contribution to economic growth and employment, but also because they display specific characteristics that differentiate them from large firms. For example, SMEs tend to be more flexible and capable of innovating, but they are bound by more stringent human resource and financial constraints, and rely on less formal managerial processes and systems, including MC and PM ones (see, e.g. Ghobadian and Gallear, 1997; Nandan, 2010). Nonetheless, they still need to foster behaviors that are consistent with the company’s aims and values, create sufficient organizational alignment and understand and improve performance (Smith and Smith, 2007) – the very tasks that are often enabled by formal control systems in large firms (Kaplan and Norton, 2008).

SMEs’ success also depends on their capacity to balance short- and long-term objectives, internal and external focus, control and creativity, and the implementation of the current strategy and the development of a new one (Simons, 2000; Frow et al., 2005; Tuomela, 2005; Henri, 2006; Mundy, 2010; Spekle et al., 2017). However, various authors have emphasized that SMEs’ underdeveloped management processes and capabilities as well as lack of resources prevent them from achieving such balance (Garengo et al., 2005; Fuller-Love, 2006; Smith and Smith, 2007). Moreover, although empirical studies have described various frameworks and high-level processes, what these organizations can actually do to measure and manage performance remains poorly understood (Ates et al., 2013; Bianchi et al., 2013).

SMEs’ economic relevance and specific characteristics, the importance of balancing different organizational demands and the dynamic interplay of formal and informal systems make these organizations an ideal context to investigate fundamental issues in PM and management theory and practice, such as how technical and social controls can be combined (Bititci, 2015). This study therefore sets out to understand how SMEs can use multiple control systems to effectively measure and manage their performance.

Specifically, we conducted an in-depth, case-based study of MC and PM practices at a small-sized Dutch company – Superior Manufacturing Group Europe (SMGE). Data were collected over a period of 17 months, using multiple data sources and data elicitation methods, including interviews and participant observations. Given the importance of balancing both technical and social forms of control in SMEs, the conceptual framework employed in this study is the levers of control (LOC), which argues that organizational performance is managed through a set of four systems (beliefs, boundary, diagnostic and interactive), and which has at its core the notion of balance (Simons, 1991, 1995, 2000). Specifically, Simons (1995) argued that, to achieve balance, organizations should use the four systems simultaneously, as they are interdependent and complementary. Because of its explanatory power, the LOC framework has played a prominent role in studies of MC and PM in management accounting research (Tuomela, 2005; Widener, 2007; Mundy, 2010; Henri, 2006; Kruis et al., 2016) and in operations management (e.g. Koufteros et al., 2014), but rarely in the context of SMEs (Heinicke et al., 2016; Massaro et al., 2017).

This research makes four main contributions to theory and practice. First, it provides an insight into how MC and PM systems are used in an SME and how these systems can help this type of firms deal with challenges specific to them. Second, it explores interdependences among the four LOC and identifies specific practices through which different LOC are jointly used to manage organizational tensions. Third, by employing the LOC framework as a theoretical perspective, the study responds to recent calls for integrating both technical and social aspects of managing performance. This is particularly important for the study of MC and PM in SMEs, as formal and informal, technical and social, and organizational and managerial issues are closely intertwined in such organizations. Finally, although many previous studies have focused on the tension between two LOC – diagnostic and interactive – this study extends this conversation to consider all four LOCs and elucidate the interactions between them.
The paper is organized as follows. The foundational literature and the conceptual framework of the study are presented in the next section. Subsequently, we describe the research design and methods employed. These are followed by the findings. The discussion of contributions, the limitations of the study and the suggested avenues for further research are presented in the final section.

**Literature review**

**Research on MC and PM**

MC and PM are fundamental organizational processes, which have attracted the attention of academics and practitioners alike in the fields of operations management, management accounting and strategy (Franco-Santos *et al.*, 2007; Melnyk *et al.*, 2014). MC has been defined as “the formalized routines and procedures that use information to maintain or alter patterns in organizational activity” (Simons, 1991, p. 49); PM as a formal process, which aims to obtain, analyze and express information about an aspect of a process, an activity or a person (Micheli and Mari, 2014). Thus, both MC and PM are processes related to the acquisition and use of information that aim to guide and influence behaviors and actions. In some cases, PM systems have been characterized as a type of MC system (Koufteros *et al.*, 2014).

Research on MC and PM has progressively shifted from the design and implementation of MC and PM systems to their uses and impacts (Widener, 2007; Pavlov and Bourne, 2011; Bititci *et al.*, 2012; Melnyk *et al.*, 2014). Empirical studies have found that effects can be positive, for example, in terms of employee engagement and performance improvement (Franco-Santos *et al.*, 2012); however, they depend on the way these systems are used and how they interact with each other (Henri, 2006; Pavlov and Bourne, 2011), as in the case of budgeting and PM systems (Malmi, 2001; Wiersma, 2009). Moreover, scholars have recognized that PM systems need to be seen as both technical and social mechanisms and have called for more research into the relationship between PM systems, organizational culture and management style (Bititci *et al.*, 2012).

In terms of empirical contexts, many studies have focused on large private firms, often to assess the design, implementation and use of popular frameworks such as the Balanced Scorecard (BSC) (Kaplan and Norton, 1992, 2008; Malina and Selto, 2001; Micheli and Mura, 2017). Some have concentrated on SMEs and argued that scorecards can be equally useful in small and large companies (Monk, 2000; Laitinen, 2002; Gumbus and Lussier, 2006; Giannopoulos *et al.*, 2013; Basuony, 2014). For example, Basuony (2014) suggests that the main benefit of using the BSC in SMEs is to promote a focus on the company’s vision and strategic objectives, as well as initiating discussion about underlying assumptions of the business model. However, this study stops short of explaining in sufficient depth how SMEs could actually use this kind of PM system. Similarly, Gumbus and Lussier (2006) conduct three case studies, but do not provide a clear depiction of how the BSC was implemented and used. Wiesner *et al.* (2007) explore the adoption of high performance management practices – for instance in relation to PM, compensation and strategic planning – in SMEs, but do not explain how these should be tailored to the specific needs of these firms. On the basis of research carried out in 15 small Finnish companies in financial distress, Hakola (2010) suggests that the BSC can be used as a comprehensive tool to recognize and improve the most critical performance aspects during the process of reorganization, thus increasing the companies’ chances of survival. This is in line with the case study undertaken by Bianchi *et al.* (2013), which concludes that adequate design and use of PM systems can lead to the identification of strategic resources driving performance and sustainable development in SMEs. On the other hand, neither study appears to take sufficiently into account the specific features of small and medium-sized enterprises. Indeed, the majority of articles that have investigated the role of MC and PM in SMEs have been either descriptive or prescriptive, but not sufficiently explanatory. In particular, they have rarely discussed in depth how specific
characteristics of SMEs could enable or hinder the introduction of PM systems, and why they might require an approach and a set of practices different from the ones adopted in large firms. In the next section, we first review SMEs' most significant characteristics and then introduce the LOC as the main theoretical framework for this study.

Characteristics of SMEs' management practices
SMEs are critically important contributors to economic growth through innovation, job creation and formation of new industries. For example, within the EU-27, SMEs account for about 60 percent of the GDP, while their contribution to value added and share of employment are 58 and 67 percent, respectively (European Commission, 2013). SMEs are often successful, thanks to their flexibility, dynamism and capacity to innovate (Laforet, 2013). However, about half of start-ups fail in the first five years and many struggle to grow, partly because of SMEs' limitations in terms of both managerial abilities and resources (Fuller-Love, 2006). Indeed, SMEs tend to be severely constrained in terms of management, manpower and financial resources; struggle to formulate coherent strategies and plans; focus on short-term priorities, which in turn trigger reactive behaviors and a fire-fighting mentality; and often rely on tacit knowledge that is rarely shared (Garengo et al., 2005; Smith and Smith, 2007; Brem et al., 2008; Bianchi et al., 2013; Ates et al., 2013; Saunila, 2016).

The lack of advanced management practices and tools has been identified as a major barrier to growth (Leach and Kenny, 2000) and a key cause for SMEs' failure (Fuller-Love, 2006; Ates et al., 2013). In this context, research indicates that MC and PM systems could play a considerable role in supporting the development of SMEs and in helping them pursue both short- and long-term goals (Garengo and Bititci, 2007). However, these systems, while promoting organizational alignment, should still enable SMEs to retain their agility, flexibility and capacity to adapt to changing circumstances (Andersen et al., 2003).

The LOC framework
To study the role of MC and PM in the context of SMEs, which have to balance multiple business demands without relying on either overly formal or excessively informal controls, we adopt the LOC framework, originally developed by Simons (1991, 1995, 2000). We do so, as this framework makes it possible to consider different MC systems jointly, to identify their main roles in an organization and to examine in depth the concept of “balance” (for similar but alternative approaches, see Marginson, 2002; Ferreira and Otley, 2009; Merchant and Van der Stede, 2012).

Simons (1995) suggested that four types of MC systems – beliefs, boundary, diagnostic and interactive – work together to benefit a firm. Beliefs systems, such as mission and vision statements, communicate and enhance core values that should inspire employees for creative and entrepreneurial thinking and guide them in their search for opportunities and new solutions. Beliefs systems are needed to ensure clear guiding principles that define the patterns of desirable behavior in relation to the core elements of an organization (e.g. values, purpose and direction), the level of the organization the organization strives for, and the expectations for managing internal and external relationships (Simons, 1995).

While beliefs systems embody an underlying positive attitude and thinking, boundary systems follow a somewhat opposite logic. These systems, which include codes of conduct, threshold targets and risk registers, communicate which behaviors are considered undesirable, which minimum levels of performance must be ensured, and set the limits to the search for new opportunities.

Diagnostic control systems aim to promote the attainment of goals. For example, budgets and performance indicators can be utilized diagnostically to help organizations monitor their processes and results, to track progress toward goals and to establish whether the desired levels of performance are being achieved (Tuomela, 2005; Koufteros et al., 2014).
Diagnostic systems provide feedback and act as a mechanistic type of control (Pavlov and Bourne, 2011) that is aligned with the notion of “single loop” in organizational learning (Argyris and Schon, 1978).

While performing an essential function, diagnostic control systems, with their focus on understanding and monitoring current performance, can also discourage managers from pursuing innovation in their processes, products or services (Henri, 2006). Therefore, various scholars have emphasized the importance of interactive control systems that are used to initiate strategic dialogue, identify opportunities and trigger change (Bisbe and Otley, 2004; Marginson et al., 2010; Pavlov and Bourne, 2011). Managers use indicators, targets and plans interactively by involving themselves in the decisions of subordinates, focusing employees’ attention on critical areas and actively engaging in dialogue and discussion (Simons, 1991; Widener, 2007), thus activating a feedforward mechanism (Pavlov and Bourne, 2011) aligned with a more organic type of control and “double loop” learning (Argyris and Schon, 1978). When used interactively, MC and PM systems can help develop and renew strategy (Abernethy and Brownell, 1999; Simons, 1991; Mundy, 2010), rather than simply support the implementation of the existing one, as in the case of diagnostic control. The main artifacts, roles and intended outcomes of the four LOCs are summarized in Table I (see also Marginson, 2002 and Mundy, 2010).

Although different, the four levers are interdependent and appear to have a more positive effect on performance when considered together than when developed and utilized separately (Widener, 2007; Bellora and Gunther, 2011; Speklé et al., 2017). For instance, empirical studies suggest that, if effectively deployed, beliefs systems positively influence the other three, as they provide broad guidance and help motivate individuals to improve current performance and search for opportunities (Widener, 2007). Also, interactions between levers should be carefully considered as they may otherwise create tensions between ensuring the achievement of an organization’s goals (the traditional “controlling use”), and the search for new possibilities and ways to improve and innovate current practices (the so-called “enabling use”) (Simons, 1995; Chenhall and Morris, 1995; Bisbe and Otley, 2004; Henri, 2006; Marginson et al., 2010; Mundy, 2010; Kruis et al., 2016).

<table>
<thead>
<tr>
<th>Levers of control</th>
<th>Main artifacts</th>
<th>Roles</th>
<th>Intended outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beliefs systems</td>
<td>Mission, vision and value and credo statements</td>
<td>To communicate and strengthen core values and priorities related to different stakeholders and financial success</td>
<td>Shared vision and values; the workforce is inspired, motivated and committed</td>
</tr>
<tr>
<td>Boundary systems</td>
<td>Codes of conduct, risk registers and indicators, threshold targets, activities explicitly categorized as off-limit, existence and enforcement of sanctions</td>
<td>To limit behaviors, identify minimum levels of performance, and to communicate risks that must be avoided or minimized</td>
<td>Employees operate within defined constraints</td>
</tr>
<tr>
<td>Diagnostic systems</td>
<td>Budgets, performance indicators and targets, rewards</td>
<td>To monitor and evaluate the progress toward goals, and indicate corrective actions</td>
<td>Achievement of critical success factors; implementation of deliberate strategy</td>
</tr>
<tr>
<td>Interactive systems</td>
<td>Strategic plans, scenarios, performance indicators and targets</td>
<td>To trigger face-to-face discussion and dialogue about performance; to continuously examine and challenge the assumptions underpinning the current strategy; and to allow for a new strategy to emerge</td>
<td>Managers and employees identify strategic uncertainties and opportunities, develop an alternative strategy and modify existing MC and PM tools</td>
</tr>
</tbody>
</table>

Table I
Main artifacts, roles and intended outcomes of the levers of control
For example, Mundy (2010) found that managers achieve balance between predictable goal achievement and creativity through the simultaneous use of MC systems in controlling and enabling ways. Importantly, balance does not imply equal weighting for the four levers, as decisions over which ones to emphasize depend on strategic and contextual factors. Also, balance can mean different things to different organizations and can be achieved in diverse ways. Indeed, in their study of MC practices, Kruis et al. (2016) identified four distinct patterns of balance, which appear suitable to different strategic intents (vigilance, exploitation, responsiveness, and stability).

While insightful, extant research has two main limitations. First, while some authors have empirically investigated how balance can be achieved effectively, they have not examined how balance can evolve over time and be open to managerial intervention, as strategic objectives and environmental conditions change (Mundy, 2010; Speklé et al., 2017; Kruis et al., 2016). An exception is Sandelin’s (2008) exploration of a relatively small, but growing, entrepreneurial owner-managed company. Results suggest that an effective combination of controls depends not merely on contingent factors (lifecycle phase, history, managerial preferences and institutional environment), but also on the managerial response to those contingencies. The author suggests that future research should be carried out on the management challenge of balancing conflicting interests and demands as well as understanding the very act of balancing (Sandelin, 2008).

Second, while some authors have considered the entire LOC framework into account (Tuomela, 2005; Widener, 2007; Mundy, 2010; Kruis et al., 2016), many have chosen to focus only on the diagnostic and interactive systems (see, e.g. Abernethy and Brownell, 1999; Bisebe and Otley, 2004; Henri, 2006). While this is potentially appropriate in the context of large firms, neglecting the roles of beliefs and boundary systems can be problematic in the case of SMEs, as the latter are likely to have fewer formal MC and PM tools that are used in diagnostic and interactive systems (e.g. strategic plans, budgets, performance targets and indicators and financial rewards). Instead, values and vision (often communicated by the owner-manager) and the explicit identification of requirements and constraints are likely to be present and to play significant roles.

In sum, on the basis of the literature reviewed, it is clear that understanding the interplay between MC and PM systems is important, and that the LOC framework constitutes a suitable theoretical lens for studying this phenomenon, especially in the context of SMEs. However, despite findings drawn in past research, the question of how SMEs measure and manage performance remains open. In particular: how are MC and PM systems used to deal with SMEs’ specific management challenges? How do SMEs utilize different MC and PM systems simultaneously to manage the identified organizational tensions?

**Research design and methods**

This study adopts an action research approach based on a single qualitative case study. The need for qualitative work on SMEs has long been recognized (Garengo et al., 2005; Ahrens and Chapman, 2006). In particular, case studies are especially appropriate for answering the “how” type of questions and gaining insight into complex phenomena in their natural settings (Voss et al., 2002). This aligns with the focus of this research on understanding how the interactions between different control systems enable SMEs to manage performance effectively. Although case studies were originally employed for theory building (Eisenhardt, 1989), their use extends beyond pure induction and includes theory elaboration (Ketokivi and Choi, 2014), making them appropriate for this study, which draws on an existing but underdeveloped theoretical base.

The study was conducted using an action research method (Westbrook, 1995), which involved a series of interventions in the case organization over an extended period of time. This allowed the researchers to adopt multiple roles and collect data through various
sources, such as semi-structured interviews, participant observations and documents. The intensity of researcher involvement required by the action research approach and the focus on gaining an in-depth understanding of the tensions involved in managing performance in SMEs limited the sample to a single case. This is consistent with the literature, which has relied on single case studies for maximizing learning about similarly complex phenomena where prior knowledge was limited (e.g. Naor et al., 2015; Pellinen et al., 2016; Beer and Micheli, 2017).

Case selection and company profile

This study was conducted in 2014-2015 in SMGE, a small-sized company located in Barendrecht, the Netherlands. The company was initially approached due to its reputation for achieving and sustaining positive financial results and for its approach to measuring and managing performance. Garengo and Bititci (2007) suggest that the development and use of PM systems in SMEs is shaped by four factors: the governance structure, the strategy or business model, the management information systems and the organizational culture and management style. The review of these factors obtained during initial visits at SMGE suggested that the number and diversity of control systems within the company provided a rich context for studying the tensions inherent in managing performance in SMEs. Moreover, the management team was willing not only to share information, but also to develop their approach further.

SMGE operates on the basis of a business-to-business model, selling commercial and industrial floor matting for professional use. The company was established in 2000 as a subsidiary of Superior Manufacturing Group Inc., a privately held company headquartered in Chicago, USA. By 2014, SMGE had grown its customer base to over 1,000 customers located in more than 40 countries, employed 27 people and generated revenues of €10 million with an EBIT of 20 percent. The company was run by a management team consisting of the managing director, financial controller, marketing manager, sales manager, operations manager and knowledge manager.

Despite being a subsidiary, SMGE operates as an independent “managerial company” (Garengo and Bititci, 2007), with a separate business model and a top management team that has the freedom to decide on strategy, set objectives and run the operations, providing the parent company with a short summary of business results. Under the leadership of this team, SMGE developed a complex set of mechanisms for managing performance. In the early years, the managing director emphasized the core values of continuous improvement and knowledge sharing, which, coupled with a strong interest in the needs of end users, led to the development of a culture of learning and innovation. This was supported by implementing appropriate systems and changing the physical design of the office. The company developed an electronic repository of documents and required all sales representatives to upload their customer-related notes onto the system in order to grow the knowledge base and support knowledge sharing. The physical layout of the premises was also changed, eschewing traditional offices in favor of a design that enabled spontaneous interactions between individuals.

However, following initial success, in 2008 the managing director recognized the need for a more structured approach to understanding and managing the drivers of performance, and in 2010 the BSC was introduced. Although technically a formal PM system derived from strategy (Kaplan and Norton, 2008), the BSC was developed in a bottom-up fashion, reflecting the collaborative and collegiate culture of the company. Each employee was consulted about the objectives within the four perspectives of the BSC, and the process was driven by questions such as: what is the unique value that we offer to our clients? How do we differentiate from our competitors? How can our success be sustained and increased?

Once in place, the BSC was used in multiple ways. As a PM system, it supported the budgeting process and provided ongoing information about the key dimensions Use of MC and PM systems in SMEs
of performance. More importantly, however, the BSC formed the basis for the weekly Monday Morning Management Meeting (the so-called “4M”) attended by the entire management team and chaired by the managing director. The purpose of the 4M was to review the performance of the previous week and to coordinate the activities for the current week. Each manager prepared and delivered a report, which was structured around the objectives of the BSC. The presentation of reports was followed by opening up a free-flowing dialogue and collective reflection on the performance information. Although the data were available in the reports, the 4M was considered to be indispensable for sharing perspectives, making sense of the operational and market reality, and informing subsequent actions. In this sense, the BSC provided the structure for an ongoing conversation about performance.

Data collection
The research involved an in-depth participative immersion in the research site, which took more than 17 months, resulted in approximately 62 hours spent in the company, and was conducted in two phases. The aim of the first phase (January-April 2014) was to examine the existing management practices and provide the basis for designing a subsequent intervention. At this stage, data were collected through semi-structured individual and group interviews with the management team and through non-participative observation of the management team meeting, resulting in a total of 26.5 hours of data collection spread over six non-consecutive days. All managers were interviewed more than once for two main reasons: first, new topics and questions emerged during data collection and required further investigation; second, repeated conversations helped verify the insights drawn from previous interviews. This iterative approach aided reflection on the part of the researchers and enabled the introduction, questioning and refinement of the research protocol, as multiple rounds of interviews supported the development and adjustment of the initial theoretical framework. In particular, repeated mentions of the relationships between formal control systems and the company strategy, culture and capacity to innovate led to a broader scope of research than originally envisaged.

This phase concluded with a case study that was written and presented to the management team in September 2014. The case raised a number of questions about performance management at SMGE, and the company decided to conduct a formal in-depth review of the existing PM system.

The second phase of the study (September 2014-May 2015) involved designing and carrying out a number of interventions. In this phase, the first author facilitated 11 workshop sessions for a total of 32 hours with the management team, focusing on the identification of strategic objectives and the corresponding indicators, targets and initiatives, and helped develop the company’s strategy map (see Appendices 1 and 2). The data were collected through participant observation, interviews and documents, which included internal files, e-mail correspondence and multiple reports and studies conducted at SMGE in the past.

Data analysis
A dynamic interplay between collection, analysis and interpretation of data characterized the entire research process, as initial insights were continuously discussed and validated with the management team. In practice, the interviews were transcribed and the documents and notes from each meeting and session were organized to enable analysis. The main analytical procedures involved coding the text from multiple sources into the categories of an a priori theoretical coding framework (Miles and Huberman, 1994; King, 2004) grounded in the literature on the LOC and summarized in Table I. Throughout the coding process, the codes were iteratively compared against the main artifacts, roles and intended outcomes of the four LOC identified in the literature. However, in order to examine the tensions between
the LOC and thus enable theoretical development, the coding process also allowed new categories to emerge through the traditional combination of open and axial coding (Strauss and Corbin, 1998).

The long time spent with the company, the participative nature of the research design and the diversity of data collection methods improved the reliability and trustworthiness of data analysis in several ways. First, the data were triangulated across multiple sources and methods to reduce bias (Denzin, 1978). Second, the emergent findings were continuously validated through conversations with the respondents. For example, the case study written at the end of the first phase was presented to all members of the management team and their feedback informed subsequent analysis. Finally, the deep involvement in the decision making of the case company allowed the researchers to improve their knowledge of SMGE, its organizational culture and dynamics, thus building an interpretive base for analyzing the findings.

Findings
We begin this section by providing an overview of MC and PM systems at SMGE using the LOC framework. Subsequently, we present two in-depth illustrations of how these systems are used: the first examines the use of MC and PM systems at the management team level; the second one describes how these systems are utilized to empower and direct employees in the sales department. Lastly, we show how MC and PM systems are used at SMGE to address the management challenges inherent in SMEs and balance various organizational tensions.

MC and PM at SMGE
Beliefs systems are extensively used at SMGE. The company strives to be the leader in its industry and expresses its main organizational values as: focus on quality, superior service, exceptional buying experience and customer support. Strong values also guide the way individuals are expected to manage external and internal relationships. Externally, the company wants to engage and provide a personal touch to the relationship with the client. Internally, the main values include open and sharing culture, low power distance, sharing success and failures. As the managing director stated: “employees feel safe to speak up. If there are problems, people talk about it. This is the culture where people are not afraid to stick out their neck and express what happens within the team.”

The managing director embodies and constantly emphasizes these core values, as they are a fundamental part of a beliefs system that defines and guides desired behavior. Managers also engage with this system informally: while eating or traveling together, they mention the core values, relate them to current targets and sometimes discuss the potential need for changes. In these conversations, the core values take concrete shape, become part of the company’s folklore and get disseminated through storytelling.

Boundary systems are in place mainly to ensure a minimum level of performance in relation to specific activities. For example, requirements include: maintaining on-time shipment (delivery within 24 hours after ordering); responding to clients’ complaints within 24 hours and offering a solution within five days; and responding to clients’ questions and inquiries within the same day. These requirements are used consistently and service excellence is regarded as a major driver of success.

In terms of diagnostic systems, until 2008 SMGE had relied on budgeting and a set of performance indicators to track and report its performance. Key performance indicators, utilized weekly or monthly, were mainly financial (e.g. liquidity, profitability and inventory turnover) and operational (e.g. backlogs). Over time, to ensure real-time monitoring of performance, screens showing real-time and cumulative sales data were installed throughout the offices and corridors.

Following the introduction of the BSC in 2010 – accompanied by the strategy map in 2015 – interactive control systems have gained more prominence. While budgets and a list
of selected KPIs are still utilized to monitor and control performance, the use of the BSC at 4Ms helps structure the dialogue around the organization’s performance and its main drivers. Even more, as the managing director argued: “the introduction of the BSC created focus on those [strategic] elements […] that are easy to remember and easy to create excitement on. In this way BSC created magic.”

Although the managers tend to receive most performance data before the meetings, the discussion of what is included in the BSC enables them to consider different perspectives and to discuss relevant aspects in relation to the company’s processes, results, clients and markets. Performance information is also considered from the perspective of the core values of the business (and the derived rules that are in place) to help interpret the events, but also to give meaning to core values and boundaries themselves, i.e. to probe, strengthen and share them.

At the company level, it is also possible to notice how different levers interact. For instance, SMGE has very well-developed information systems, and uses ERP and CRM systems to support the diagnostic use of budgets and KPIs. However, the display of information also helps to increase employees’ involvement and commitment and trigger a more interactive use of information. Similarly, since SMGE decided to work through distributors rather than to sell directly to the end customers, it has limited contact with the users of its products, thus potentially limiting learning and subsequently undermining the core value of innovation embodied in its beliefs system. However, this is counteracted by a boundary system rule that stipulates that all clients who spend more than €50,000 must be visited at least once a year.

**Simultaneous use of different LOC**

We now provide two salient illustrations of how SMGE used different LOCs simultaneously to manage organizational performance. In order to denote the interactions between the individual LOCs we observed, throughout this section we abbreviate the terms beliefs (Bf), boundary (B), diagnostic (D) and interactive (I) and show the effect of one system on another in pairs; for example, the effect of the beliefs system on the diagnostic system is shown as BfD.

**Management meeting.** As described earlier, a core element in managing performance at SMGE is the 4M, which is chaired by the managing director and where individual managers’ reports are followed by a dialogue that enables collective sensemaking. The reports usually contain financial and non-financial indicators, and the discussion revolves around topics such as sales (to investigate important inquiries that can turn into orders), supply chain and logistics (to organize the incoming supplies), operations (to handle the incoming containers), knowledge management and marketing (to reuse the existing materials in order to supply relevant information to clients) as well as significant orders and the resolutions of customer complaints. The 4M creates a positive environment where the tensions between multiple LOCs are managed simultaneously with an eye toward improving organizational performance.

For example, the yearly budget, which informs all managers’ reports, is consistent with SMGE’s vision, mission and long-term strategic objectives, contains specific targets, and supports the implementation of the intended strategy (BfD). Similarly, the 4M emphasizes the discussion of budget variance figures, particularly about sales, but also about inventory levels and costs. The whole management team is involved in the discussion about the reported results and events, and therefore the issues are considered from different functional perspectives, taking into account market trends and feedback received from clients (DI): “the estimation of the sales for the future period is really a team effort as it is about knowing your market, following trends in different geographical segments we operate in and also knowing dynamics of exchange rates” (managing director).

The dialogue that emerges during the 4M is structured around the strategic objectives defined in the BSC: “the information that we share during the meeting and that we
[Voluntarily] report to the parent company is all organized around the BSC that provides the structure (managing director). The insights from this discussion become the input for the formulation of performance targets for the following year (ID).

During the discussion, individual views are communicated, exchanged and compared until a shared understanding is reached. This dialogue gives fresh meaning to the core values, verifying, sharing and strengthening or adjusting them (IBf). When considering specific issues in the reports, the assumed relationships between strategic objectives and performance standards are challenged to test if they still hold (e.g. is fast delivery really what makes our clients happy and is that likely to continue in the future?) (IB). Superior achievements are also observed from the perspective of the company’s chosen strategic approach, thus reinforcing the core values (DBf). Likewise, during the 4M, the existing rules are challenged and put into broader perspective. Very often stories about past events are told, which give more meaning to the rules that are in place (IB): “everyone knows the story of IKEA. […] by responding to their repeated complaints and requests to solve their problem we came up with a new product line” (managing director).

The performance data are also used to support non-negotiable standards. For example, one of the promises the company gives to its clients is that the orders will be delivered within 24 hours. For that reason, backlogs are monitored and discussed daily and reported regularly (BD). Backlog figure variances and the reasons behind them are also discussed with the aim to learn from the mistakes and make improvements in the future (DI): “during the meetings we celebrate success. We also talk about mistakes, and we try to see how we can learn from them” (managing director).

Through the dialogue at the 4M, the company’s core values are used to inspire and guide the search for new opportunities. By doing that, the beliefs system is used as a powerful LOC. The ideas exchanged at the 4M are themselves clearly influenced by SMGE’s values: take care of the client, discuss and cooperate in order to find the best solutions, share both success and failure and learn from them (BfI). Moreover, the nature of the dialogue at the 4M is shaped by the open and low power culture, where people are encouraged to voice their views and are not afraid to question, to share, or to suggest alternative perspectives (BfI): “I have trust in my management team. If someone makes a comment or observation, they probably had a good reason to do so. So I really try to understand what they meant and why they said it” (managing director).

Managing the sales department. The simultaneous use of multiple LOCs is also evident in the way SMGE manages its sales department. Every Monday, following the 4M, a sales meeting takes place. It is chaired by the sales manager and attended by all sales representatives and the managing director. As in the case of the 4M, the sales representatives prepare a short report showing the previous week’s data, including most significant orders, promising prospects and significant issues (e.g. complaints, inquiries). An open discussion follows, where the sales representatives provide valuable insights into the clients’ behaviors and needs, and the senior managers help the sales force interpret the data in the broader context of strategic uncertainties and critical success factors. The interactive use of the data stimulates the process of exchange, collective thinking and reflection through dialogue. While managers implicitly use the BSC to define the structure for both reporting and discussion and to determine priorities in specific situations, it is the continuous and open dialogue in the meeting that enables collective learning (DI). An interesting feature of the sales meeting is that rules, such as responding to clients’ inquiries, are treated as thresholds and monitored through the reporting system (BD).

The simultaneous use of multiple LOCs extends beyond the meeting. For example, the core values have been translated into specific rules that explicitly guide the activities and communicate what is expected from the sales representatives (BfB). The managing director
emphasized this point: “if we want to be become a leader, we need to start to behave like a leader. We want to convey the perception of quality through marketing materials and every form of communication. Further, the product finishing, packaging, labeling, handling and shipping are all important aspects that should reflect an image of quality.”

However, one of the issues in managing the sales force is that in the short run the responsibilities of sales representatives can be conflicting. For example, they need to follow the professional norms and the company procedures while at the same time fit the culture that emphasizes creativity. SMGE recognizes this challenge and looks to recruit people who can balance this tradeoff (BBf and Bf B): “For this task you need a particular profile of a person that can comply with the rules, but is also comfortable with having freedom. In other words, we are looking for a sales person who can accept the rules, but can, within these rules, perform a task in a different, creative way. These ways are different because we are really international and for each region the sales person is a native and knows exactly how to talk to a client from that region and what makes them tick. It is all about setting boundaries but also giving freedom” (sales manager).

Similarly, sales representatives are responsible for achieving targets while being required to contribute to the attainment of strategic goals formulated in the BSC such as improve the quality of experience and responsiveness to inquiries and leads. Such tensions are also managed by the simultaneous use of multiple LOCs. For example, in order to contribute to the improvement of the quality of service, sales representatives are expected to respond promptly to the client’s complaint. Dealing with the complaint, however, slows down the progress toward reaching the sales targets. In order to manage this tension, SMGE explicitly incorporates its core value of customer centricity in the design of the sales report, requiring the sales representatives to report both on the generation of new business and on the quality of service in the same document (Bf D): “I use several control elements […]. Some of them are very explicit, such as a sales target or a request to respond to the complaint within 24 hours. Further, sales personnel has to report, prepare a weekly highlight every week. The headings [i.e. topics] for the weekly highlight are set by me and they completely correspond to the BSC objectives. So, when they report they need to touch on new clients, new products, responsiveness to inquiries and leads, and improving the quality experience of clients. So they are all aware that this is what matters and they need to contribute to it” (sales manager).

Moreover, the exploration of the reasons for complaints gives insight into a particular use of the product. When this information is shared and discussed back at SMGE, it stimulates the search for product improvement and new product development (DI).

The responsiveness to inquiries and leads objective, similarly to the investigation of complaints, is considered critically important for exploring new possibilities and developing new products. Sales representatives have a responsibility to record and provide specific information about the products requested by clients. In most cases, this is a routine task that takes time away from current sales. As in the example above, SMGE uses the beliefs system to motivate sales representatives to evaluate customer inquiries and discuss them with management in an interactive process. SMGE clearly communicates the core values of learning and innovation and emphasizes that the company needs to capture the leads and to gather and record all information about end users’ experience. Every remark that end users make about the use and performance of SMGE products is considered valuable, and sales representatives are encouraged to record them both in their individual reports and in the CRM system, and to discuss them with management and in the sales meeting (Bf I). As a result of this process, some of those inquiries have become sources of product improvements and innovations as well as sales to new clients.

In sum, the illustrations of the simultaneous use of multiple LOCs reported above provide an insight into how SMGE uses all four LOCs to manage performance. Successful strategy
implementation at SMGE is ensured by maintaining the focus on the budget and the KPIs supported by a powerful use of the beliefs system. At the same time, a focus on innovation is ensured through an interactive use of the BSC within clearly defined standards of service and behavior, particularly with respect to customer relations. Using the BSC as a core PM system, the management continuously balances the diagnostic and the interactive systems. The former is enacted through the use of budgeting and performance reporting; the latter through continuous dialogue (formal and informal) around BSC objectives. The beliefs system, in the form of strong values and a clear vision, is used to inspire and guide the search for new opportunities so that the company can remain responsive to strategic uncertainties. Formal reports, the 4M and the sales meetings provide the structure, and the behavioral norms and performance thresholds set the boundaries within which interactive control can usefully operate. In other words, the diagnostic, beliefs and boundary systems set an appropriate context for the necessary dialogue to evolve. It is this dialogue, supported by an open and low power culture, that enables the interactive use of the BSC and ensures that the scorecard triggers creativity and innovation. The identified interactions between multiple LOCs are graphically represented in Figure 1, and the practices employed by the company to enable these interactions are reported in Table II, where the intersection of the rows with those in the columns refers to the effect of the former on the latter.

**Discussion and conclusions**

This study provides an in-depth examination of MC and PM practices in a small firm. In doing so, it addresses recent calls for research that takes into account the specific characteristics and challenges of SMEs and considers both technical and social aspects of PM (Smith and Bititci, 2017).

Our findings contribute to the advancement of PM and management theory and practice in several ways. First, this study provides an in-depth illustration of the range and diversity of conflicting goals and tensions that organizations can manage through the simultaneous

![Figure 1. The interactions between the four levers of control in SMGE](image-url)
<table>
<thead>
<tr>
<th>Beliefs systems</th>
<th>Boundary systems</th>
<th>Diagnostic systems</th>
<th>Interactive systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beliefs systems</td>
<td>Core values are translated into specific rules for sales representatives. Recruiting individuals who can observe norms and procedures while engaging in innovative and proactive behaviors.</td>
<td>Vision, mission and long-term strategic objectives inform the budget. The core value of customer centricity is incorporated into the design of the sales report.</td>
<td>The company values and culture shape the ideas and the flow of the discussion on strategic planning in the management meeting. Sales representatives are encouraged to capture and discuss all customer enquiries to support the core values of learning and innovation. Boundaries set the limits for the search of new opportunities. Some solutions are not accepted if they do not fit with promises made to clients.</td>
</tr>
<tr>
<td>Boundary systems</td>
<td>Recruiting people who can observe norms and procedures while engaging in innovative and proactive behaviors.</td>
<td>Performance against defined standards is regularly monitored and evaluated. Rules (e.g. to respond to clients' inquiries) are converted into performance targets and monitored through the reporting system.</td>
<td></td>
</tr>
<tr>
<td>Diagnostic systems</td>
<td>Achievement of performance objectives is interpreted as reinforcing the core values. Performance data are used to support non-negotiable standards.</td>
<td>Insights from discussions are used to formulate the following year's performance targets.</td>
<td>Market trends and customer data are fed into open debates. Budget variance figures used for future looking discussions. BSC categories are used to structure the discussion at the meetings and enable sharing, reflection and learning. Collective exploration of the reasons for complaints promotes search for product improvement and innovation.</td>
</tr>
<tr>
<td>Interactive systems</td>
<td>Continuous dialogue is used to give fresh meaning to the core values. During meetings the assumed relationships between strategic objectives and non-negotiable standards are challenged to test if they still hold. Discussions and stories are used to give meaning to rules.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
use of multiple MC and PM systems. In the case company, the interdependence and complementarity of these systems helped the organization balance several competing perspectives: short- and long-term focus, predictable goal achievement (strategy implementation) and search for new opportunities (development of a new strategy), internal and external focus, and control and creativity. Specifically, in SMGE, different time horizons were simultaneously taken into account by viewing current events in connection with longer term strategic objectives, market trends and risks. To implement the current strategy and at the same time allow a new one to emerge, KPIs and budgets were often used to understand and drive performance while consideration of customer feedback and continuous monitoring of strategic uncertainties encouraged the creation of future strategy. Focus on both internal and external factors was ensured by complementing the attention paid to internal practices and processes with the use of externally focused core values and service quality targets. A clear focus on innovation, an open and sharing culture, and specific recruitment practices on one hand and concentration on achieving expected results and controlling deviations from targets on the other enabled the company to balance creativity and control. Importantly, balancing different organizational demands also helped trigger both single and double loops of learning, so that the continuous improvement of processes and products was interspersed with the introduction of new offerings.

Second, this study extends prior work by using the LOC framework to identify and organize a set of practices, summarized in Table II, that the case organization deployed to measure and manage its performance. Indeed, earlier research has made important advances in developing frameworks to describe multiple control systems (Simons, 1995), theorizing the importance of maintaining balance (Henri, 2006) and identifying various contingencies affecting such balance (Mundy, 2010). However, this discussion has often remained on a relatively high level, and the question of which specific practices organizations could actually deploy to balance multiple control systems in pursuit of better performance is still poorly understood (Ates et al., 2013; Bianchi et al., 2013). The set of practices identified at SMGE illustrates the interdependencies between the four LOCs and provides the foundation for future research into specific managerial actions required for ensuring the balance between different forms of control.

Third, the findings highlight specific aspects related to PM and management in SMEs. While our results resonate with those of previous studies conducted in large companies (see, e.g. Henri, 2006; Mundy, 2010; Koufteros et al., 2014; Kruis et al., 2016), SMGE, similarly to most SMEs, adopted less technical and formal mechanisms to manage performance (Garengo et al., 2005) and the managing director played a critical role. Indeed, a strong reliance on the beliefs and interactive systems was a prominent feature of its approach. The use of such systems emerged as a fundamental – and arguably less costly and controversial than more formal mechanisms such as rewards (Franco-Santos et al., 2012) – means to motivate employees and stimulate commitment, creativity and innovation. However, it was not simply a matter of communicating and reinforcing the company values, nor was this a task undertaken exclusively by the managing director. The positive role played by the beliefs and interactive systems was ensured by two mutually reinforcing factors: the beliefs systems were constantly enacted by all senior managers and aligned with the management style and the organizational culture; beliefs and interactive systems were mobilized through dialogue and discussion and were tightly linked to the boundary and diagnostic systems (see Table II). Indeed, the technical and social controls at SMGE were continuously and purposefully connected.

The results of this study thus demonstrate that ensuring an active interconnection between all four LOCs is essential. As such, the study extends the current debates on technical and social controls (Smith and Bititci, 2017) by moving beyond the identification of structural relationships between individual systems (e.g. Heinicke et al., 2016) and instead
showing that the deployment of a full range of controls is a multifaceted, ongoing and purposeful process requiring active involvement on the part of the management. Moreover, by providing an in-depth empirical analysis of the simultaneous use of all four LOC (Widener, 2007; Mundy, 2010), this research extends earlier work (see, e.g. Abernethy and Brownell, 1999; Bisbe and Otley, 2004; Henri, 2006) that tended to consider two levers only (typically diagnostic and interactive). In the case organization studied here, the achievement of balance and the positive interplay between different MC and PM systems could only be explained by considering all four levers together.

This research also has implications for PM and management practice. In particular, the identified complementarity of the four levers indicates that managers should consider them together when designing or reviewing their approach to managing performance. This is particularly salient in the context of SMEs where less formal controls, such as beliefs and interactive systems, are likely to play a key role (Heinicke et al., 2016), and where balance should be ensured despite the lack of managerial processes and capabilities. More generally, various combinations of control systems should depend on the particular strategic and contextual circumstances in which an organization operates (Kruis et al., 2016). Importantly, as these combinations are difficult to imitate (Mundy, 2010), they may constitute a unique organizational capability and thus a source of sustainable competitive advantage.

This research also speaks to the body of work that considers the design, implementation and use of PM systems (Bourne et al., 2000; Kaplan and Norton, 2008; Ferreira and Otley, 2009; Melnyk et al., 2014). More specifically, we found that the BSC “created magic” – as stated by SMGE’s managing director – when there was a clear connection between objectives and KPIs, and when performance information was utilized interactively, rather than simply diagnostically. In other words, employees’ engagement with the PM system increased when data were used to contextualize and shape future actions and longer term objectives, rather than simply to understand current and past performance. When used interactively, the BSC also functioned as a framework for organizing verbal and mental thought processes for examining the relevance of basic assumptions of the company’s objectives, strategy and operations. Moreover, echoing earlier research (Martinez et al., 2010), PM meetings seemed to be the central element that linked performance information and action and enabled the BSC to be closely tied to the belief and boundary systems, thus making it a fundamental tool that employees would extensively use.

This study has limitations, which provide potential avenues for future research. The findings are based on a single case study, and, despite this not being unusual in PM and MC research (see, e.g. Tuomela, 2005; Mundy, 2010; Naor et al., 2015; Pellinen et al., 2016; Beer and Micheli, 2017), larger qualitative as well as quantitative studies are needed to understand whether the interplay between the four LOC is similar in other contexts. Such studies could also compare various configurations of LOCs and evaluate their effects on performance. While we expect similar patterns in both SMEs and large firms, future quantitative research could consider salient organizational characteristics (e.g. managerial capabilities, formalization of organizational processes and size) and examine their effect on companies’ capacity to use multiple LOCs simultaneously and thus to measure and manage performance more effectively. Moreover, SMGE was selected on the basis of its positive financial performance and its deployment of various MC and PM practices. Contexts characterized by less sophisticated approaches may display different dynamics and less successful outcomes.

Finally, the prominent role of the managing director, and more broadly the management team at SMGE, suggests that another interesting direction for future studies is the examination of the relationship between control systems and leadership style, especially in SMEs where the founder or managing director plays a fundamental role (Smith and Smith, 2007). It has already been indicated that a manager’s strategic awareness (Hannon and Atherton, 1998), background
and experience (Lucas et al., 2013) and individual temporal orientation (Sternad, 2014) have an impact on his or her leadership style and, ultimately, the company's performance. It is possible that firms balance their control systems differently depending on the type of leadership style (Collier, 2005; Speklé et al., 2017) and future studies could provide a useful insight into the nature of this relationship.

References


Use of MC and PM systems in SMEs


Appendix 1

Date | Participants | Topic | Topics discussed
--- | --- | --- | ---
September 22, 2014 | Management team | Customer perspective, 1st session | The review of the existing Balanced Scorecard (BSC) started with the customer perspective. Conversations aimed at clarifying the value proposition of the company; the objectives included in the customer perspective motivated managers to share their perspectives and views. The researcher moderated and guided the discussion, but the managing director played a central role, as he was genuinely interested to hear and understand department managers’ perspectives and to see how these could complement each other. The main objectives (retain existing and acquire new customers) and their drivers (on-time shipping, customization, quality of the experience, and capacity to respond to inquiries and leads) were identified. These drivers encapsulate the promise SMGE makes to its clients and differentiate SMGE from its competitors. Some indicators were also discussed.

October 3, 2014 | Management team | Customer perspective, 2nd session | Continued discussion of the objectives in the customer perspective, mainly from a strategic point of view, e.g. who are the right customers? What does the company stand for? Members of the management team were encouraged to openly present their ideas, to validate them and to contribute to building a shared understanding. Sharing views also helped break barriers between different departments and positions, especially supply chain, marketing, sales and finance. This process, facilitated by the BSC review, was also perceived to enable a shift from an operational to a strategic focus.

October 10, 2014 | Management team | Customer perspective, 3rd session | The customer perspective was finalized, also by checking that the identified objectives would lead to sustainable growth. This question led to a discussion about the company’s values and beliefs: understanding the needs of clients; developing loyalty and trust; showing interest in clients; and keeping a personal touch. The conversation then evolved into a dialogue on whether the rules (boundaries) in place were clear and justified. The meeting concluded by examining the objectives in the internal business perspective.

November 21, 2014 | Management team | Internal perspective, 1st session | Critically important activities and operations were discussed. The achievement of required performance in these areas is ensured by rules that are instituted and known to everybody as well as elements of the beliefs system. The rationale behind the main rules was probed. The session concluded with a discussion on the relationship between SMGE and its suppliers, and a story was told about cooperation with a Thai supplier.

February 20, 2015 | Management team | Internal perspective, 2nd session | Discussion about objectives and indicators in the internal business perspective. In this session the conversation was about the importance of quality, including what quality means to different departments and what is expected from them. A shared understanding of quality and critically important aspects that affect it was reached; these include relationships with suppliers, which could be improved through greater communication and cooperation.

March 6, 2015 | Management team | Internal perspective, 3rd session | In the third session devoted to the internal perspective, the discussion centered on the contribution of the sales and marketing departments to improving quality, as perceived by the client. Continued attention to service and product quality demonstrates the importance of this topic for the company.

Table AI. Summary of the workshops focused on the BSC review (continued)
<table>
<thead>
<tr>
<th>Date</th>
<th>Participants</th>
<th>Topic</th>
<th>Topics discussed</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 13, 2015</td>
<td>Management team</td>
<td>Internal perspective, 4th session</td>
<td>Discussions about it allowed different views to be exchanged and taken into account. The main outcomes of this process were the agreement over a shared view of quality and the identification of main responsibilities in the business. This was followed by a meeting with the MD to capture the key stories illustrating the company’s values and identity.</td>
</tr>
<tr>
<td>March 20, 2015</td>
<td>Management team</td>
<td>BSC review</td>
<td>This session was devoted to discuss the rules (boundaries) in place and verify that they are still valid and relevant. Indicators were also reviewed to ensure that they captured key aspects related to the attainment of the most important objectives. Review of the whole scorecard to verify that it encapsulates the most important objectives and that cause and effect links are present. Agreement of all managers was sought and the dialogue, which had taken place in the previous sessions, contributed to better understand the other departments' perspectives, interdependencies and challenges.</td>
</tr>
<tr>
<td>April 10, 2015</td>
<td>Two senior managers</td>
<td>BSC review</td>
<td>Review of all the existing performance measures. During the last session the focus was on how to ensure people's motivation and enthusiasm. The beliefs system was seen as critically important to motivate employees and to get their buy-in. Continuous dialogue, both formal and informal, was regarded as a key means to ensure that all employees accept and understand the importance of the core values and rules for the success of the company.</td>
</tr>
<tr>
<td>April 17, 2015</td>
<td>Management team</td>
<td>Learning and growth</td>
<td>Review of the financial perspective: objectives, measures and initiatives.</td>
</tr>
<tr>
<td>May 12, 2015</td>
<td>Management team</td>
<td>Financial perspective</td>
<td></td>
</tr>
</tbody>
</table>

Table AI.
Appendix 2

Use of MC and PM systems in SMEs

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com

Corresponding author
Biljana Pešalj can be contacted at: b pesalj hr nl
Predicting performance – a dynamic capability view

Marcus F. Hasegan
Manchester Metropolitan University, Manchester, UK
Sai Sudhakar Nudurupati
GITAM School of International Business, Gandhi Institute of Technology and Management, Visakhapatnam, India, and
Stephen J. Childe
University of Plymouth, Plymouth, UK

Abstract

Purpose – Production planning and resource allocation are ongoing issues that organisations face on a day-to-day basis. The purpose of this paper is to address these issues by developing a dynamic performance measurement system (DPMS) to effectively re-deploy manufacturing resources, thus enhancing the decision-making process in optimising performance output. The study also explores the development of dynamic capabilities through exploitation of the organisational tacit knowledge.

Design/methodology/approach – The study was conducted using six-stage action research for developing DPMS with real-time control of independent variables on the production lines to study the impact. The DPMS was developed using a hybrid approach of discrete event simulation and system dynamics by using the historical as well as live data from the action case organisation.

Findings – Through the development of DPMS and by combining the explicit and tacit knowledge, this study demonstrated an understanding of using cause and effect analysis in manufacturing systems to predict performance. Such a DPMS creates agility in decision making and significantly enhances the decision-making process under uncertainty. The research also explored how the resources can be developed and maintained into dynamic capabilities to sustain competitive advantage.

Research limitations/implications – The present study provides a starting-point for further research in other manufacturing organisations to generalise findings.

Originality/value – The originality of the DPMS model comes from the approach used to build the cause and effect analysis by exploiting the tacit knowledge and making it dynamic by adding modelling capabilities. Originality also comes from the hybrid approach used in developing the DPMS.

Keywords System dynamics, Dynamic capabilities, Discrete event simulation, Cause and effect analysis, Dynamic performance measurement system, Predictive performance

Paper type Research paper

Introduction

To fully understand the root causes leading to excellent or poor performance in manufacturing has been a debate in both operations management and performance measurement literature (Bititci and Nudurupati, 2002; Suwingnjo et al., 2000). Exploring the relationships between various input factors or resources affecting manufacturing performance is important for operations strategy and operations decision-making (Silvestro and Lustrato, 2014; Tan and Platts, 2005, 2009). Measures such as overall equipment effectiveness (OEE), utilisation and throughput are predominantly used to improve manufacturing productivity (Huang et al., 2003). The vast majority of decision-making tools such as Aggregate Production Planning tools were developed and tested using mainly explicit knowledge (Jamalnia and Feili, 2013). While there is abundant research on how to measure them (Liu et al., 2004; Ahmed and Sahinidis, 1998; Silver et al., 1998), there is comparatively little research on identifying the inputs and resources (root causes) that influence them (Jeong and Phillips, 2001). These bundles of resources can constitute a basis for competitive advantage and for the development of dynamic capabilities (Rothaermel, 2013; Teece et al., 1997).

While the purpose of performance measurement is to measure the effectiveness and efficiency of actions (Neely et al., 1995), the majority of information is based on static and...
explicit knowledge with emphasis on tangible assets (Neely, 2005). Given the challenges of fast changing business trends (Bititci et al., 2012) and the advent of information technology (Harrington et al., 2011; Harrison and Van Hoek, 2011), organisations need to be fast in making decisions, flexible in planning their resources and proactive in addressing the changing strategies to retain competitive advantage going beyond the exploitation of the measurable operational knowledge (Neely, 2005; Melnyk et al., 2014). While Neely (2005) argues the need for dynamic performance measurement systems (DPMSs), Melnyk et al. (2014) confirm the lack of such models to support businesses in implementing and monitoring their strategies. Neely (2005) also argues the need for measuring both tangible and intangible assets thus drawing the focus on capturing both explicit and tacit knowledge (Anand et al., 2009) in understanding the performance of the business. Improved understanding – particularly as production conditions change – is vital to measure performance and to manage resources. Few researchers have explored the usefulness of modelling approaches that could complement performance measurement systems (PMSs) by identifying and predicting future performance thus making them more dynamic (Warren, 2008; Santos et al., 2002; Bititci, 1996).

The current focus of trends in modelling manufacturing performance literature is on discrete-event simulation (DES) (Negahban and Smith, 2014; Chopra and Meindl, 2007). To overcome some of the DES limitations and to capture the dynamics of the manufacturing processes, Santos et al. (2002) attempted to develop PMSs using system dynamics (SD). This approach models the relationships between possible causes and effects and develops dynamic decision support tools (Venkateswaran and Son, 2007; Rabelo et al., 2005). There are multiple challenges and difficulties for practitioners to find and develop sound methods or tools for capturing the existent knowledge inside the organisations (Sanchez, 2001). While explicit knowledge has been fruitfully exploited (Jamalnia and Feili, 2013; Bar-Yam, 2006; Greasley, 2005; Brailsford and Hilton, 2001), the tacit knowledge existent inside the organisations has not yet been used in dynamic modelling.

The overall aim of this research is to build a DPMS for modelling the key input resources and capabilities to not only evaluate the current manufacturing performance but also to predict future performance. With DPMS, organisations are capable of making fast resource re-deployment decisions and can constantly review the variable inputs that need continuous improvement (CI). The purpose of this paper is twofold. First, by developing and demonstrating the use of a modelling approach based on cause and effect analysis in real-time to predict future performance and contribute to the existing body of literature (Melnik et al., 2014; Unahahbokha et al., 2007; Suwingnjo et al., 2000; Bititci, 1996; Neely et al., 1995). The research also fills the gap in knowledge by complementing the existing models that use DES and SD using explicit knowledge (Peteraf, 1993; Barney, 1991) through the use of CI for identifying the critical variables that can be improved, thus creating a dynamic system. Although such models are complex, once developed they are difficult to copy, thus giving organisations a competitive advantage (Peteraf, 1993; Barney, 1991). The human resources and intellectual property of an organisation are more valuable in combination than separately (Anand et al., 2009). Second, to solve a real-world problem by developing a DPMS that models the relationships between a number of key variables driven by operational knowledge (metrics) and tacit knowledge generating production output. Practitioners can model the variables and redeploy their resources to achieve optimum performance. The detailed objectives of this study are:

- to understand the causal relationship between the factors that influence manufacturing output using explicit and tacit knowledge;
- to build the DPMS using DES and SD to model the factors that impact manufacturing output; and
- to test the DPMS and its impact on the organisation.
A UK tyre-manufacturing factory was selected as a case organisation in this study for two reasons: first, because planning and scheduling is generally complex in tyre manufacturing (Tabucanon and Petchratanaporn, 1991). Second because the tyre industry demand is unpredictable (Sull and Escobari, 2004). In addition, the selected organisation was facing performance issues and struggling to identify the root causes of issues in a fast-moving environment while meeting day-to-day needs. The planning lacked dynamism and decisions were taken later after the system's performance has been assessed. There were issues with their resource planning which had a knock-on effect on the production. The main PMS they used was OEE, however in a very reactive approach. Decisions were delayed and issues were escalated to the higher management. There was a genuine need to understand various resources deployment issues that would impact their bottom line and proactively manage their production. In order to improve the decision-making process, there was a need for a decision support tool that would help managers deploy the resources in right place at the right time to improve effectiveness and efficiency.

Action research was chosen as the main methodology for achieving the overall aim. The data were collected in five different ways: personal observation; from the shop floor production control and data capturing system (both historic and current); leading and participating in multiple workshops and CI events; semi-structured interviews with employees in both formal and informal meetings; and from documents such as meeting minutes, performance measurement white papers and communication documents. Based on the data collected a simulation model was developed, tested and used in the organisation to support decision-making, and the findings were evaluated. This study used the dynamic capabilities perspective as a means to theorise some of our findings to contribute to the existing literature (Rothaermel, 2013; Crook et al., 2008; Eisenhardt and Martin, 2000; Teece et al., 1997; Wernerfelt, 1984, 1995).

The paper is structured as follows: it starts with the literature review and presents a detailed evaluation on predictive performance measurement. It also reviews the existing literature on discrete event simulation (DES) and SD and on hybrid approaches in manufacturing that add dynamic aspects to PMSs. It continues with a section that reviews the dynamic capabilities and the link between the company’s tacit and explicit knowledge and the potential development of these dynamic capabilities. This is followed by the presentation of the overall research design, which includes the data collection and analysis. The findings are then presented using the six-stage action research framework implemented in the case organisation. The next section presents the findings and finally finishes off with conclusions and further work.

**Literature review**

Performance measurement and management is a mature field with a number of contributions for designing, implementing, using performance measures for decision-making and improvement (Bourne et al., 2000; Bititci, 1996; Neely et al., 1995). However, the majority of the existing models and frameworks were built on measuring only tangible assets/resources and they are static in nature, thus making them less relevant when the strategy changes (Melnyk et al., 2014; Bititci et al., 2012; Franco-Santos et al., 2012; Neely, 2005). Hence understanding the factors influencing performance outcomes in a business or manufacturing environment is crucial for evaluating or predicting performance against their strategy (Bititci and Nudurupati, 2002; Suwingnjo et al., 2000). A few researchers have argued that modelling approaches when used with PMSs will create dynamism in evaluating or predicting performance outcomes (Warren, 2008; Santos et al., 2002; Bititci, 1996).

Hence, in order to achieve the overall aim, it is useful to explore existing literature on predictive performance measurement as well as modelling approaches, and these are presented in the following sections. The most common modelling approaches used in the
literature are DES, SD, and hybrid approaches, which are discussed here. As the emphasis is on measuring tangible and intangible assets (hence the need to capture both tacit and explicit knowledge), the dynamic capabilities perspective is discussed.

**Predictive performance measurement**

Time has been described as both a source of competitive advantage and the fundamental measure of manufacturing performance (Stalk, 1988). Other than time, there are multiple factors creating, impacting and transforming the competitive advantage (Porter, 1985). Understanding the sources of competitive advantage has been a constant preoccupation for researchers (Barney, 1991; Porter, 1985; Rumelt, 1984). Multiple challenges have been raised and some researchers claim that competitive advantage cannot be sustained in dynamic, rapidly changing markets (Eisenhardt and Martin, 2000). However, they acknowledge that some companies are more agile, able to change quickly and more willing to change, thus gaining an advantage. Barney et al. (2001) suggest that the value of dynamic capabilities must be evaluated in a market context and if the context changes these capabilities will no longer be valuable. Franco-Santos et al. (2012) argue that PMSs’ effects impact people’s behaviour, organisational capabilities and organisational performance in a changing environment. Melnyk et al. (2014) highlighted the pace of environmental change as an ongoing issue and PMS were criticised as unreliable in practice for not being sufficiently dynamic. In other words, while strategy changes rapidly, the PMSs were much slower in response, resulting in tension, misunderstood effects and eventually losing their relevance. Hence there is a need for tactical tools to create organisation specific competencies in a changing environment.

Fast decision-making in manufacturing is mainly linked to organisational agility (Yusuf et al., 1999). Agility is defined as “the capability of surviving and prospering in a competitive environment of continuous and unpredictable change by reacting quickly and effectively to changing markets, driven by customer-designed products and services” (Gunasekaran, 2001). However, there are impediments to building an agile, innovative enterprise. An organisation may lack sufficient knowledge about itself to know where and when change is needed. Christensen and Overdorf (2000) as well as Winter (2003) have pointed out that important capabilities are often embedded in the less-visible and background processes that support decisions relating to areas such as investment and resource reallocation. Tangible resources tend to depreciate over time (Porter, 1985) so the sole exploitation of those alone may not sustain competitive advantage in long term.

Barr (2014) highlights the need for more research on PMSs including intangible goals and assets. She also calls for people to be engaged in using the measures which affect a company’s strategic and tactical levels. Bititci and Nudurupati (2002) argue that the closed-loop control system is necessary to continuously monitor the performance of processes to identify and improve the parts of the process. They also argue that performance indicators need to be designed and the relationships need to be modelled in order to sustain CI (Kaizen) in the outcomes. Suwingnjo et al. (2000) has explored a number of tools including cognitive mapping, cause and effect diagrams and tree diagrams and developed quantitative models for PMSs to quantify the effect of predictive measures on the top line performance. This work is further extended and enhanced by Sarkis (2003) to incorporate various feedback loops using the analytical network process to predict top line performance. Similarly, Tseng (2010) explored a multi-criteria evaluation approach to use balanced scorecard and determine the dependence, analytical network processing, as well as interactive relationships (decision-making trial and evaluation). These findings suggest that with the advancement of simulation and modelling techniques, there is a greater opportunity to model the relationships of various resource inputs and capabilities to study their impact on top line performance and throughput.
Discrete event simulation (DES)

DES quantitatively represents the real world, simulates its dynamics on an event-by-event basis and generates a detailed performance report (Law and Kelton, 2000). In the early days, DES was used to pre-test flow layouts in fully automated manufacturing systems (Wu and Wysk, 1989) or in a low-volume, mixed model Just-in-Time assembly system (Carlson and Yao, 1992). Welgama and Mills (1995) used a simulation approach to address design problems faced by a chemical organisation, changing from a traditional to a Just-in-Time system, considering alternative designs for the Just-in-Time system. DES is widely used and an increasingly popular method for studying the design and operations of manufacturing systems (Rabelo et al., 2005; Kleijnen, 1995). In DES, state variables change only at discrete points in time, called “event times” (Brailsford and Hilton, 2001).

For DES, accurate historical data are needed to produce statistically relevant results. Exploring the dynamics of the DES, Chong et al. (2003) used simulation based scheduling for dynamic discrete manufacturing by using off-line simulation experiments in manufacturing environments that are subject to disturbances based on scheduling approaches. However, DES lacks dynamism and all the reasons for the estimations and causes of correlations cannot be deduced but must be inferred. Moreover, it does not allow the researcher to determine the stability of the system. While DES has long been a popular technique for studying industrial processes, it is also used widely for planning and evaluating design alternatives in a production process (Law and Kelton, 2000; Oakshott, 1997). Wohlgemuth et al. (2006) suggests that when simulation models are properly validated, they can be used to answer questions and suggest improvements in complex systems. Exploration of these models can significantly improve the understanding of a modelled system’s behaviour.

The difficulties of understanding complex systems are explained by Sterman (2001) as due to the stakeholders failing to understand the full range of feedback operating in the system. In summary the literature notes that DES has a short-term impact (Anand et al., 2009; Greasley, 2005) and cannot efficiently model complex systems (Rabelo et al., 2005). It is proposed that integration with SD can generate mid and long-term results (Venkateswaran and Son, 2007).

System dynamics (SD)

An increasingly popular simulation method used in production modelling is SD. SD was developed by Jay Forrester in the mid-1950s to gain a better understanding of the behaviour of complex systems. It is an analytical modelling methodology which combines both qualitative and quantitative aspects (Brailsford and Hilton, 2001). According to Hidaka (1999), SD can be applied in three thinking frameworks: current situation analysis, causal analysis and solution selection. SD assists in strategy development (Sweetser, 1999). It captures inputs in causal loops and enables the decision maker to model and compare the performance of a system over a range of alternatives (Sweetser, 1999; Kleijnen, 1995).

Sweetser (1999) suggests that SD is useful to model the system as stocks and various flows in pseudo-continuous time. However, Robinson (2004) suggests that both discrete event modelling and SD can support decision making. While some argue that DES and SD are quite separate simulation approaches (e.g. Brailsford and Hilton, 2001), others see them as complementary to one another (Morecroft and Robinson, 2005). SD creates a bridge between long-term strategy and short-term planning (Venkateswaran and Son, 2007). The challenges of quantifying the interactions between input factors were highlighted by Baines et al. (2005). They suggest that the gap between actual and predicted performance is due to models failing to incorporate key relationships such as the human impact on performance. There is also an obvious gap between production planning modelling, which operates in discrete time (Díaz-Madroño et al., 2014; Kadar and Monostori, 2001) and factory modelling which operates on a continuous basis (Hidaka, 1999). These findings suggest that there is a need for a more robust integration of DES and SD to fill this gap.
Hybrid approach

To overcome the difficulties of static modelling, since the early 2000s there has been a call for reconfigurable manufacturing systems. Mehrabi et al. (2002) suggest reconfigurable manufacturing systems as the key to future manufacturing. These systems must be rapidly designed and be able to adapt quickly to changing needs. Hence there is a need to change both discrete and continuous factors in order to be able to respond to disruptions. Santos et al. (2002) suggest that the usefulness of the integration SD into the PMSs can be described as a “better identification of the key elements of success”.

Hybrid systems are those where discrete and continuous factors co-exist (Größler et al., 2003; Huang et al., 2003; Lee et al., 2002). Qiao et al. (2003) proposed a data-driven design and simulation system to support flexible manufacturing. The model can be modified quickly to adjust manufacturing capabilities. Rabelo et al. (2005) proposed a hybrid approach to cope with modelling manufacturing systems with increasing complexity. They also emphasise the role of SD in shifting the traditional DES focus from individual decisions to policy structure. In an attempt to create a dynamic model by using only the explicit knowledge existent in the organisation such as robustness, fill rate and inventory level, Sun et al. (2012) use multiple simulation models to compare performance. Greasley (2005) attempted to use a DES and SD hybrid approach presenting a case study in a manufacturing environment, highlighting the limitations of a DES in capturing important qualitative data likely to improve the relevance in decision making. In his study, although the usefulness of DES was valuable, the performance of the company was impacted by working practices as well. That study addressed the lack of research in the integration of the tacit knowledge into dynamic modelling as well as their complexity.

Several researchers raised the importance of the possible qualitative issues occurring in making manufacturing decisions (Gregoriades and Karakostas, 2004; Größler et al., 2003; Levin and Levin, 2003). They also stressed the importance for the model to be able to stabilise the processes. Sterman (2001) argues that attempts to stabilise systems may actually destabilise them and Forester calls such a phenomena the “counterintuitive” behaviour of social systems. However with DES and SD approaches, a strong set of criteria is needed for the processes modelled to be sustained and improved (Rabelo et al., 2005). That is mainly because unexpected dynamics often lead to the tendency for interventions to be delayed or defeated (Sterman, 2001). In the vast majority of the organisations, to keep decisions within cognitive bounds, managers must often simplify processes extensively (Russo and Schoemaker, 1989).

Dynamic capabilities

Dynamic capabilities define a firm’s ability to innovate, adapt to change and improve in a way that is favourable to the customer and unfavourable to their competitors (Teece et al., 2016). Dynamic capability is defined as a “learned and stable pattern of collective activity through which the organization systematically generates and modifies its operating routines in pursuit of improved effectiveness” (Zollo and Winter, 2002, p. 340). A firm’s dynamic capability should govern how it integrates, builds and reconfigures internal and external competences to address changing business environments (Winter, 2003). To manage uncertainty, organisations must have strong dynamic capabilities (Teece et al., 2016).

Winter (2003) argues that an organisation’s ordinary capability is a high-level routine, or collection of routines that, together with its input flows, confers upon an organisation’s management a set of decision options for producing significant outputs of a particular type. Teece et al. (2016) argue that ordinary capabilities stem from the proficient employment of the firm’s human resources, assets (tangible and intangible), processes, and administrative systems, including the coordination needed to combine in-house and external resources. The strength of a firm’s ordinary capabilities is a measure of its technical fitness.
Unlike ordinary capabilities, dynamic capabilities are based on developing, carrying and exchanging information through the firm’s human capital (Hitt et al., 2001). Dynamic capabilities contrast with ordinary capabilities by being concerned with change. Capturing the tacit knowledge and creativity possessed by the shop-floor people fulfils the CI infrastructure function of bottom-up generation of process improvement ideas (Winter, 2003). Anand et al. (2009) presented a framework of CI infrastructure derived from the dynamic capabilities perspective and its underlying theory of organisational learning (Zollo and Winter, 2002).

Capabilities at an abstract level can be thought of as the ability to co-ordinate activities, learn within an organisation, and re-configure resources. Teece et al. (1997) have referred to learning as a specific type of process underlying dynamic capabilities, which is based on repetition, experimentation and identification of new opportunities. On the other side, dynamic capabilities require a longer-term focus and involve subordinating short-run cost cutting, optimisation, and other “best practices” to (longer time) innovation-enhancing strategies. Wernerfelt (1984) argues that it is necessary to view organisations in terms of resources rather than their products to have competitive advantage by creating resource position barriers. He also calls for more research identifying and developing those niche resources by combining capabilities across different divisions as well as developing structures and systems to implement such strategies for entry barriers. According to Anand et al. (2009) process improvement involves organisational learning to make changes in operating routines.

A number of papers discussed the links between capabilities, resources and routines and their contribution to the organisation agility and prosperity (Ismail and Mamat, 2012; Grant, 2003). Many treatments of agility (or flexibility) in the management literature would seem to suggest that firms should persistently seek to become agile no matter the cost, keeping options open all the time, maintaining redundancy at all times and staying in a constant state of radical transformation (Teece et al., 1997). Due to an increasing pace and complexity of business environments, organisations no longer compete on processes but the ability to continually improve processes (Teece, 2007). Apart from the identification and assessment of the technological opportunities which can successfully be done using DES (Sterman, 2000) there is also a need for the identification and mobilisation of relevant resources through the exploitation of tacit knowledge, which can be modelled using SD (Sterman, 2000). Hence there is a need for continued renewal or transformation process facilitated by a CI framework for superior organisational performance.

It is clear from the literature that DES and SD are increasingly used for dynamic modelling in the manufacturing environment. The literature also demonstrates the usefulness of studying and developing hybrid models. The role of the tacit knowledge in the development of competitive advantage as well as in the success of complex dynamic models has been previously emphasised in the literature (Sterman, 2000; Teece, 2007). Only a few studies include tacit knowledge in any of its forms into dynamic models, warranting the need for further study. Therefore, this research uses tacit knowledge in building a DPMS by using a hybrid model to evaluate the current performance and predict the future performance. The emphasis is not in developing another model per se, but the approach for combining performance measurement, modelling techniques and tacit knowledge for applying dynamism to predict future performance. The method used to collect purposive data is crucial in achieving the aim and is discussed in the next section.

Method
In order to develop and test the DPMS we needed full experimental control in manipulating the input variables to study the impact on OEE, planning patterns and production output. The solution was necessary for a real-world problem, which requires personal observation, participation, control and engagement. Hence this study adopted action research as the main strategy in achieving the objectives. Action research involves practical problem
solving (with experimental design) which has theoretical relevance (Coughlan and Coghlan, 2002; Mumford, 2001; Gill and Johnson, 1991). The approach aims at both taking action and creating knowledge or theory about that action (Coughlan and Coghlan, 2002). Gill and Johnson (1991) describe action research as a study in which researchers/practitioners, from their interventions and subsequent evaluation not only contribute to the existing knowledge but also solve the practical concerns of the people. In this study, the researchers need to participate and engage in the organisational change simultaneously studying the processes with full access to experimental design and control over the project or context (see Myres, 2009). At the same time, the researchers should also engage in academic activity to identify theoretical significance.

According to Dyer and Wilkins (1991), single in-depth case studies are richer than shallow multiple case studies. Hence a single action case was selected to spend more time and effort over a two-year period to collect in depth data and highly perseverant analysis (Jarvensivu and Tornroos, 2010). This study took place in a tyre manufacturing facility in the UK owned by one of the biggest tyre producers in the world.

Case context

Tyres are made from a range of rubber and metal components that are combined and transformed in a number of different stages. There are three main stages in the manufacturing process. The initial stage is the preparation stage in which the raw materials (natural rubber) are mixed with various chemicals resulting in uncured rubber compound. All these semi-finished products are then transformed in the building stage. The building stage transforms the semi-finished product first into a “carcass” with calendered (heat treated) ply and sidewalls. In the second stage of building, the carcass is transformed to a “green” cover with the metal belts, the tread and the spiral. Finally, at the curing stage the green cover is cured. After quality tests are passed, the finished product is ready to be delivered to the customer. This research looked at the process and the information flow management of the building stage and of the curing stage, which constitute the core areas of the business.

The manufacturer’s processes were facing severe disruptions. There were significant delays in decision making. First information was escalated hierarchically and most of the times an answer came late. The company also lacked contingency plans in case disruptions occurred. Mainly, decisions were taken based on gut feeling by the team leader or manager and different practices were used on different shifts. For the last ten years there had been the same patterns of behaviour involving a slow decision process, low productivity with an increased number of orders backlog and unfulfilled orders. The company lacked both strategic and tactical planning.

Tactically, the production planning lacked a reliable resource allocation in case of disruption which had a knock-on effect on their production. Another important issue was that some of the variables that can affect the throughput, such as waiting times in the process due to work practices or waiting times for managerial decisions, or lack of communication protocols, were barely known to the top management. Strategically, there was a genuine need to understand various input variables (both operational and tacit) that would impact the production throughput. There was a clear need for both a tactical and strategic decision support tool that would help managers with a fast and robust decision-making process they could rely on.

Data collection

One of the researchers in this study was based on the company’s site every day for two years. The data were collected in five different ways. First, through personal observation and engagement. The researcher was placed in the company for two years to facilitate the change, measure the improvement and study its impact in the organisation. Another researcher visited the organisation once every two weeks to monitor the research project.
progress throughout this two year study. Second, data came from the shop floor production control and from a robust data capturing system (daily, weekly and monthly efficiency and quality reports). For data validation and for testing the model, a number of reports were pulled out from the data systems (both historical and current data). While a majority of the reports were downloaded from the pre-built queries in the system, some reports were obtained through making requests to IT department. Third, data were gained by leading and participating in three workshops and four CI events. These workshops and events were facilitated purposefully to obtain more information for the research project and constituted as a method to capture the initial tacit knowledge from the organisation’s members. At these events, data were captured through post-it notes and flip charts, identifying bottlenecks, concerns, ideas and solutions and time and motion studies. Fourth, data were gathered through 60 semi-structured interviews with 28 participants with an overview of their profiles listed in Table I. Finally data were drawn from documents such as meeting minutes, and performance measurement and communication documents. The findings were triangulated to ensure internal validity.

**Approach for building DPMS**

The data obtained from various sources were triangulated and organised for manual analysis. Observation of data, content analysis and pattern matching were used where appropriate in the study (Yin, 2014). The study was structured on a six-stage CI process action research framework as shown in Figure 1. The first stage consisted of creating project team with members from relevant departments involved in the tyre production process (quality, operations, R&D, external contractors, senior management, planning). It also included operators and efficiency specialists as well as managers. The second stage involved interviewing and gathering the stakeholders’ requirements for identifying performance outcomes to monitor or improve. The third stage consisted of auditing the manufacturing system for gaining a better understanding of the causes and effects as well as a better understanding of the possible noises, and a cycle of CI events was kicked-off. The events, which involved all the relevant people from the shop floor, aimed to understand the complexity of the manufacturing system and to identify all the root causes affecting manufacturing performance. Causal loop diagrams were also developed and refined to facilitate the model development and also to predict future performance. The model, with variables and causal relations, was presented at workshops for more insights and cross-validation.

The fourth stage consisted of developing a simulation model to understand and evaluate the impact of resources and capabilities on the performance outcomes. As a first step, appropriate simulation and modelling software was identified to support the planning decisions identified in the previous stage as shown in Table II. Sysdea powered by strategy dynamics was selected as the best fit for this purpose. A DES was designed, developed and

<table>
<thead>
<tr>
<th>Participant job role</th>
<th>Number</th>
<th>Core responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial director</td>
<td>1</td>
<td>Managing the whole plant</td>
</tr>
<tr>
<td>Mini-factory managers</td>
<td>3</td>
<td>Daily managing the building, curing and quality plants</td>
</tr>
<tr>
<td>Efficiency leaders</td>
<td>3</td>
<td>Managing the efficiency metrics and the plant capacity</td>
</tr>
<tr>
<td>Shift managers</td>
<td>5</td>
<td>Managing a whole shift</td>
</tr>
<tr>
<td>Shift coordinators</td>
<td>5</td>
<td>Managing manufacturing lines</td>
</tr>
<tr>
<td>Team leaders</td>
<td>7</td>
<td>Managing manufacturing cells</td>
</tr>
<tr>
<td>Operational quality</td>
<td>2</td>
<td>Quality checks to comply with customer specs</td>
</tr>
<tr>
<td>Industrial quality</td>
<td>2</td>
<td>Define the quality procedures in the plant</td>
</tr>
</tbody>
</table>

Table I. Summary of profiles of selected participants in the organisation
implemented which modelled the relationships between OEE, production capacity, manning resource, machine resource and production output for this manufacturing system.

The fifth stage consisted of alignment and the testing of the model to the business needs. The model was validated and revised at the CI events in three ways. First, it was internally validated with the supply chain specialists from the company. Second it was empirically validated with historical data to verify results against the historical events. Finally, the model was tested with live data to predict results against the forecasts. The sixth stage consisted of the implementation of the model. In this respect, the staff were trained to gather data, to plot and input the data and to interpret results. Moreover, based on the desired output the staff were also trained to re-assign or re-deploy the resources in the simulation model until the desired output was achieved.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Simul8</th>
<th>Sysdea</th>
<th>Witness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to create cause and effect analysis of output variables with their associated input variables</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ability to see manufacturing performance based on historical data</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ability to manipulate some of the input variables, i.e. re-deploy resources to see the impact on future performance</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ability to see a visual representation in the form of charts, histograms, etc., for instance, to see build-up/depletion of inter-stage inventories while making resource deployment decisions</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Seamless or easier integration of the software with the existing data sources, i.e. Excel spreadsheets</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Tailored price package to suit the usage requirements</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Continuous support for development of extra tools and capabilities</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ability to develop holistic models built based on sub-models which simplifies the way data are presented</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Visibility of the positive and negative feedback loops</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Table II. Comparison of simulation and modelling software

A dynamic capability view
Findings
The core stakeholders’ requirements identified in this study were the following. First, the top management highlighted the low OEE and the struggle to schedule production. Second, they stressed the need for a system to help them with effective resource allocation as well as predicting production output. Third, they stressed the importance of a better understanding of a potential link between the input variables, stoppages, manning, methods and production volume that effect performance outcomes. Finally, the shop floor operators, in contrast, raised a concern about the waiting time for making decisions, which caused both frustration and delays and had a domino effect on the production throughput.

The core output factors agreed to be modelled in this case are OEE, inter-stage inventories and production output. After carrying out cause and effect analysis, the research identified the main factors (sometimes called variables, resources and capabilities or operational knowledge) that were affecting the performance outcomes are associated with production losses, both man- and machine-related, manning, and methods used by production operators (tacit knowledge).

The building stage was not synchronised with the curing stage due to complexity and side effects in the processes and the lack of decision visibility. The planning was centralised and every time a disruption occurred in the process, the upstream production stopped and the resources were wasted. Any disruptions or variances in building stage were absorbed by the slack capacity in the curing stage. Carcass inventory accumulated between any of the stages, for example see Figure 2. However, the buffers between 1st stage building, 2nd stage building and curing are seen as critical because these three stages account for the vast majority of the stoppages and production losses thus contributing to the backlog of customer orders. Hence these inter-stage inventories needed to be managed to prevent the production stopping because of upstream issues and to decrease the number of backlog orders.

Downstream planning involved matching the manufacturing capacity of one area (such as building) with the transforming capacity of another (say curing). Often the planners were making decisions based on a trial and error approach due to a lack of tools and techniques available to support their decision-making. The planners deploy the downstream resources arbitrarily by taking a risk margin, which is often between 10 and 30 per cent. Due to the asynchronous nature of building and curing stages, a lot of resources including people and machines were not deployed appropriately and hence losing efficiency.

The manufacturing process was slow because of multiple waiting times incurred in the decision process. Every time a disruption occurred the planners and the operator waited for a decision on resource redeployment. Moreover, there were four different shifts in every area, and some of them were faster than others as their ways of working were different, clearly highlighting the need for process standardisation across the manufacturing facility.

A model of DPMS was built in Sysdea by gathering data on static cause and effect analysis and inputting them into the modelling software to show current manufacturing performance. The aim was to facilitate the management to redeploy the resources and predict the impact on manufacturing performance. The key elements of the model are time, manning, volume, stocks, demand, product range and Kaizen points. The Kaizen points are the variables that can and will be continuously improved as part of the model. The SD feedback loops were built based on “before” and “after” scenarios highlighting the variables prone to CI. While the majority of causal links were based on standard formulas defining
their relationships, other links were defined using regression. Some relationships between stocks (i.e. man power, inter-stage inventory, availability of workforce, etc.) with other relevant variables were defined based on historical data and experience. However, these relationships are kept to minimum to reduce uncertainty in the overall model fitness.

The model’s inputs take into account four aspects: the number of machines running, the speed at which they are running, the number of products that meet the customer’s specification, and the number of people deployed to operate these machines as well as the methods used during the processes. Moreover, as soon as a model is created the feedback loop shows the possible side effects on the optimum resource re-deployment. The model also predicts the capacity of the tyre production line and is tailored for different types of machines. Because cycle time differs with every machine type, a different sub-model is needed for each line, with a bigger model integrating them all. Some of the variables feeding the availability and the productivity have a direct impact on the number of tyres built. The basic assumption of the model is that a machine builds at its maximum capacity providing there are no stoppages and it operates on a full shift with all comfort breaks covered and with no stoppages or other speed losses. However, during the operating time of the machine, there are both foreseeable and unforeseeable stoppages. The foreseeable stoppages are the ones occurring on a regular basis such as: planned absence, preventive maintenance, industrialisation, operators’ rest, change overs, routine operations, etc. The other stoppage types are subject to unpredictability and include breakdowns, quality issues, waiting for material, etc. While predictable variables have a constant value (or slightly fluctuating within control limits to incorporate natural variation), the unpredictable variables were given random values in the model following a normal distribution curve (an assumption suggested by Hussin and Hashim, 2011). These variables can be continuously improved, becoming Kaizen Points.

The model was evolved on a number of iterations based on historical data across a wide range of processes leading to different products. This not only improved the simulation model accuracy but also highlighted the reasons behind the current manufacturing performance. The model evaluated alternative supply chain designs for optimisation of resource allocation based on decisions suggested by a simulation/modelling approach. The model was also tested with historical data to quantify and visually demonstrate how the resource allocation. Figure 3 represents the main causal loop used in building the DES. The balancing loop operates when a problem is identified in the manufacturing process, thus putting the customer order into backlog, which eventually becomes an unfulfilled order. The unfulfilled orders have an impact in slowing the manufacturing process, demanding flexible decisions to be made in order to fulfill the backlog orders with priority. This delay has a negative impact on the productivity and it builds the carcass stock. The increase of the carcass inventory put pressure on the curing process which needs to consume these carcasses. By not consuming the carcasses it delays the orders, adding orders in the backlog. The reinforcement loop highlights the actions that produce the decrease in building carcasses and increase in cured carcasses, thus releasing curing capacity (the manufacturing bottleneck).

The model was tested and validated in three phases on a tyre production line. In the first phase, a pilot manufacturing process was chosen and historical data were plotted into the model. The project team checked the accuracy of the model and the eventual inconsistencies against the factual data. With minor tweaks the modelled scenario approached the scenario from past real-life events. In the second phase the team gathered field (real-time) data and plotted it into the model, analysing the response of the model in respect to the resources allocated as well as the inter-stage inventory and their effect on manufacturing performance. The model was tested for two weeks and the data were recorded on a daily basis. Apart from the stoppages and all other explicit data, the number of people assigned to the building machines was recorded as well as the impact of the methods used during the process. In other words, it recorded the total amount of time spent on the building machine. Next, the
observed data values in the real-life are manually inputted in the simulation model software, and compared against the predicted values from the system. Once again the modelled scenario approached the real-life scenario.

In the third phase, with support from the researchers, the senior managers used the model in their decision-making and re-allocated the resources based on the model suggestions to verify improvements in the manufacturing performance. That is, if the predicted production output was lower than the desired output, the capacities and resources were increased or redeployed in building and/or curing stages. This phase not only tested the accuracy and usefulness of the model but also built senior management confidence in the model, which increased other stakeholders’ interest in the model. The most important finding in this stage showed that by using the DPMS the company became more agile by reducing its decision times, increasing its flexibility in planning and becoming faster in re-deploying resources in case of disruptions.

The model is open, dynamic and real-time and some input variables (Kaizen Points) are included in the company’s CI strategy. These features gave strength and sustainability to the model enabling the company to focus on the improvement of the manufacturing processes. The CI method gave the company the opportunity to bring in the dormant tacit knowledge and exploit it.

Results
The case company is operating in a dynamic and complex manufacturing environment with challenges from a multitude of variables that could affect manufacturing or production performance. Without a DPMS, the decision process was extremely slow as in case of disruptions, the team leader’s or manager’s decisions were needed. In addition, the production planners were deploying downstream resources arbitrarily using their gut feeling and by taking risk and following trial and error approach. Downstream planning involved matching the manufacturing capacity of one mini-factory with the transforming capacity of another.

The development of DPMS helps the management in their decision-making process of resource allocation and in predicting the optimum resource allocation and the performance
of the internal supply chain. The DPMS dynamically models a multi-echelon, multi-product system with periodic ordering and evaluates alternative resource allocation with respect to the system’s maximum capacity. Production planners are now using the DPMS in making informed decisions, which also empowered them and shortened the decision-making process. By entering all the information on input measures (such as manning, machine availability, breakdowns, etc.) the planners are able to predict manufacturing performance (OEE, inter-stage inventories, production output, labour utilisation, etc.) more accurately. It gives them power to run what-if scenarios to see its impact on performance and redeploy (change, increase or decrease) resources until the desired outcome is reached before actually implementing changes in live production.

The building and the curing operators are focussing on the Kaizen points (changeovers and quality checks) following newly established standard operating procedures. These Kaizen points are constantly reviewed (every 6 months) and re-adjusted for the model to constantly reflect the company’s operating routines.

The DPMS offers both the benefit of holistic planning of the production and the opportunity to individually model production lines separately in order to have a quantifiable impact on the whole. Hence it has both a strategic as well as tactical planning usability. In the presented model it is used for the scenario based-approach in which the variables are generated statistically using data on either forecast stoppage percentages or normally distributed data where machine stoppages occur. The assumption is that breakdowns or quality issues occur following a normal distribution curve. The Kaizen points are the operating routines that can be improved and constantly reviewed. These variables represent the operators’ ways of working (changeovers, quality checks, waiting times, information flows) and represent the tacit knowledge inside the organisation. In addition to problem solving, the Kaizen points promote the learning of individuals through the interaction and the interpretation of a given situation.

The DPMS was used in Kaizen events and other CI projects in different ways. First, it enabled some of the action plans to be modelled to see any potential improvement before implementing these actions (sensing). Second it enabled the improvement teams to define the Kaizen points and see whether the input variables are operating between the limits upper control limit (UCL) and lower control limit (LCL) (3sd from the mean) to verify whether the processes are stable (diagnosing). Finally by reducing the decision-making time (team leaders and planners avoided escalation), it increased the agility in planning and it increased the agility in optimum resource deployment in case of disruptions (reconfiguring).

The model was used by the team leaders for optimum resource allocation for achieving the number of acceptable units required by the customer. When there were changes in planned resources, the model suggested resource re-allocation and contributed to strategic decisions on resource re-deployment in order to achieve the planned performance. In addition, the model was used to see how the inventory is likely to accumulate or deplete in a defined time period based on their existing plans and number of sizes in production. This helped planners when redeploying resources to keep inventory in the optimum limits.

The method used in designing and implementing the model through CI cycles helped the researchers and the company to identify which processes can be constantly adjusted and streamlined (quality checks and changeovers) transforming the operating routines into dynamic capabilities. These processes are used in the model as critical variables or Kaizen points. Prior to using DPMS, decisions were always made by senior management based on their gut feeling (intuition) and deployed to lower levels. People on the shop floor who were involved with production on a day-to-day basis had little input into planning decisions. As demonstrated in Figures 4 and 5, with the use of DPMS, the central and bureaucratic decision-making has evolved into local, informed and empowered decision-making. The middle management and supervisors on the shop floor were able to make planning decisions locally based on the information provided by the model. This resulted in a cultural shift driving improvement projects.
Discussion and management implications

The lack of clarification of relationship between the widespread resources in the manufacturing process and their impact on performance outcomes is amongst the main reasons why manufacturers are far away from developing efficient manufacturing systems (Soloukdar, 2012). Moreover, the failure to capture all relevant feedbacks in a complex system is one of the main causes of failure in modelling complex systems (Sterman, 2000). These concerns can be impediments in building agile and fast decision-making systems. Moreover, it also prevents the company developing a competitive advantage against the competitors (Teece et al., 1997; Rothaermel, 2013). Hence researchers call for a dynamic model, which can cover all relevant aspects (i.e. inputs, outputs, inter-stage inventory, manning resources) (Teece et al., 2016; Christensen and Overdorf, 2000; Winter, 2003).

Cause and effect analysis has been a popular method for building many CI projects. While it is a common tool used to understand complex systems, it is a static representation of relationships between inputs and outputs. The significant development in this paper is to take these static relationships and make them more dynamic by quantifying the relationships, modelling them using both the explicit and tacit knowledge existent in the business and creating a CI framework. While the need for a DPMSs was identified in literature (Melnyk et al., 2014; Bititci et al., 2012; Franco-Santos et al., 2012; Neely, 2005; Bititci and Nudurupati, 2002; Suwingnjo et al., 2000), the drive to develop such system was obtained from modelling approaches (Warren, 2008; Santos et al., 2002; Bititci, 1996).

Silver (1992) suggest that quantitative models are useful for decision making if they represent the problem realistically and permit some of the “usual givens” to be treated as decision variables. Although using output measures such as OEE as well as the effects of inputs and their dependence is not entirely new (see Tseng, 2010; Sarkis, 2003; Suwingnjo et al., 2000) few studies explored, tested and validated the input factors and their impact on OEE and production output in a real and dynamic manufacturing environment using DES and SD. Hence the emphasis of this research is in developing a visual and intuitive approach (as demonstrated in Figure 3) for deploying manufacturing resources and inventory policies in an effective and efficient manner to influence and predict performance. The explicit and tacit knowledge used in the development of this dynamic model in which the cause and effect relationships are clearly defined is likely to give a dynamic capability to the organisation (Esterby-Smith et al., 2009) which can be transformed into a temporary competitive advantage to the organisation that implements it (Rothaermel, 2013; Barney, 1991).
The proposed DPMS model is a hybrid between DES and SD to model forecasted data to give projections for future planning. The hybrid model overcomes some of the concerns raised by researchers with regards to the limitations of the DES and SD used in isolation (SD cannot map batch production and lacks accuracy in modelling while DES in isolation does not suggest feedback loops) (Jovanoski et al., 2013). In order to sustain the simulation model, a CI framework is proposed, which is a robust method for capturing the existent tacit knowledge. In addition, the model and CI framework incorporates sensing (identify changes and opportunities), diagnosing (develop new ways of responding to changes) and reconfiguring (reorganise existent operating routines) capabilities (Gonzalez and Martins, 2016).

The originality of the DPMS model comes from the approach used to build the cause and effect analysis by exploiting the tacit knowledge and making it dynamic by adding modelling capabilities. The DPMS can capture variables, named Kaizen points, which can be improved through Kaizen events with shop floor teams participating in the continuous analysis and development of the input variables. The risk of not capturing important but not so obvious variables in complex manufacturing systems is also mitigated. Moreover, the operational routines are continuously adjusted and enhanced regularly for sustaining superior performance (Helfat et al., 2007; Teece, 2007; Winter, 2003).

The new capability was organised in such a way for the company to capture its value. The relevant input variables were identified as CI variables thus offering the company valuable tools for continually enhancing those processes. By highlighting the potential of some of the manufacturing stages the model created common expectations, behaviours, and goals hence it enabled the changes in capacity and the ability to reduce cycle time of all activities.

Decision making in planning the design and operations of manufacturing processes is often based on several factors, of which some of them are uncertain. In this action research, uncertainties arose in capacities or constraints of manning or machine resource allocation, conditions such as breakdowns, potential quality issues and optimum inventory levels. On top of all these explicit and quantifiable variables there are variables influenced by the operators ways of working which sometimes can heavily influence the manufacturing throughput. All these uncertainties affect the performance of the manufacturing system, including its service levels and delivery lead-times, which in turn affect the business competitive environment (Liu et al., 2004).

The current research brought some practical implications. First, it brought value by capturing key tacit and explicit resources, their dependence on performance as well as a tool for optimum deployment. Second, it brought a rare intangible resource, a blend of DES and SD developed through employing a CI technique covering all the relevant dynamic aspects of manufacturing. Third, through the development of the CI method it created a dynamic capability, which encompasses both the strategic and tactical views of company. Finally, the model was embedded in the organisational culture through CI processes organised to continuously capture value. In essence, an organisation can model their dynamic capability to enable them to gain competitive advantage in a constantly changing environment.

The model can also help practitioners to categorise the losses as chronic and sporadic as suggested by Jonsson and Lesshammar (1999), define kaizen points and stabilise and improve losses which are normally governed by uncertainty. The current research mitigated the uncertainty risk by organising Kaizen Events to control some of the input variables, which varied historically exceeding the UCLs or the LCLs. The method is preferred to the main alternative design of experiments (DOE) due to time and cost constraints (Kleijnen, 1995). Moreover, the tactical tool is aimed to be used shift by shift whereas DOE or ANOVA might be too difficult to be used by the operational staff.

**Conclusion**

The DPMS acted as a predictive measurement system and supported the implementation of manufacturing strategy by controlling input variables such as resource allocation,
disruptions, inventory, changeovers, etc. (Nudurupati et al., 2011; Suwingnjo et al., 2000). It also inputs explicit variables built on existing data as well as critical variables (Kaizen points) that can be improved over time. Controlling the critical variables is highly likely to control uncertainty and disruption. Hence, the measurement system not only works as passive control, but is also used as CI (Jonsson and Lesshammar, 1999). The model has strong momentum for initiating CI projects as it highlights such opportunities.

The paper has presented a DPMS which has modelled multi-stage production, with multiple machine families with different behaviours. The model has demonstrated its usefulness in exploring and controlling the predictable and some of the unpredictable variables in the manufacturing system by suggesting a day-to-day resource allocation. The use of visual and predictive aspects of DPMS has transformed the company’s central and bureaucratic decision-making into local, informed and empowered decision-making. In the broad context, this paper contributes to operations management literature by demonstrating in practice that a DPMS can be created by combining performance measurement models with modelling approaches. As Melnyk et al. (2014) highlighted, when the strategy changes in the organisation, the model needs to be revised by updating new output measures and input measures in the light of new strategy. Through this research the use of static relationships in cause and effect analysis is enhanced to dynamic relationships to evaluate and predict performance in a new way, thus contributing to CI and performance measurement literature (Suwingnjo et al., 2000; Sarkis, 2003; Tseng, 2010).

In practice, the DPMS enabled the company’s tangible (explicit) and intangible (tacit) resources to transform into dynamic capabilities (manufacturing flexibility and agile resource re-deployment), which are likely to create competitive advantage. The DPMS is a tool that can be used to understand, manage and enhance the manufacturing performance through the following capabilities: sensing to observe changes and opportunities and initiate CI initiatives; diagnosing to identify the problems, route causes, stabilise and to improve uncertain processes; and finally reconfiguring to redeploy resources and routines and enhance the dynamics of the model.

Despite the fact that the model developed in the current research is specific and unique to this organisation, the approach (i.e. the way DPMS was developed by defining Kaizen points and by building it through multiple CI cycles) is transferrable to other batch production make-to-order manufacturing businesses where the benefits will be replicated. The case has also demonstrated the positive impact the DPMS had on the organisation’s decision-making process and on increasing agility in planning and predicting outcomes at both strategic and tactical levels.

Although the findings were based on one action case, the authors are confident that the findings are generalisable to other similar manufacturing companies due to the analytical nature of the solution presented (Yin, 2014). The important element of the DPMS is its accuracy of prediction, which depends on the ability to define relationships between different variables, i.e., while most of the relationships are derived by using formulas (which are accurate and reliable), some have to be defined based on correlation and regression (where the accuracy could fall down). If the manufacturing environment has more of these uncertain relationships then it will affect the accuracy of DPMS and hence its suitability will be limited. Hence in future more of these studies should be performed in different manufacturing settings such as fast moving consumer goods, make-to-stock or high-shelf life inventories under various settings to extend and strengthen this research. It is also necessary that future researchers should focus more on objectives methods such as DES and SD, particularly on hybrid approaches when predicting business performance. Objective methods such as multi-criteria decision analysis could be used with SD in enhancing the effectiveness of selecting measures during design and implementation PMS while taking input from various stakeholders (Santos et al., 2002). These analytical
approaches limit subjectivity, ambiguity and conflict between measures thus improving the effectiveness of measuring and managing performance. Further studies are also required to expand the scope of evaluating and predicting manufacturing performance to the full business, taking other aspects into consideration.

References


Further reading


Waters, C.D.J. (1992), Inventory Control and Management, John Wiley & Sons, Chichester.

Corresponding author
Sai Sudhakar Nudurupati can be contacted at: sudhakar.nudurupati@gitam.edu

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
A decision theory perspective on complexity in performance measurement and management

Anthony Alexander, Maneesh Kumar and Helen Walker

Department of Logistics and Operations Management, Cardiff Business School, Cardiff, UK

Abstract

Purpose – The purpose of this paper is to apply the aspects of decision theory (DT) to performance measurement and management (PMM), thereby enabling the theoretical elaboration of volatility, uncertainty, complexity and ambiguity in the business environment, which are identified as barriers to effective PMM.

Design/methodology/approach – A review of decision theory and PMM literature establishes the Cynefin framework as the basis for extending the performance alignment matrix. Case research with seven companies explores the relationship between two concepts under-examined in the performance alignment matrix – internal dominant logic (DL) as the attribute of organisational culture affecting decision making and the external environment – in line with the concept of alignment or fit in PMM. A focus area is PMM related to sustainable operations and sustainable supply chain management.

Findings – Alignment between DL, external environment and PMM is found, as are instances of misalignment. The Cynefin framework offers a deeper theoretical explanation about the nature of this alignment. Other findings consider the nature of organisational ownership on DL.

Research limitations/implications – The cases are exploratory not exhaustive, and limited in number. Organisations showing contested logic were excluded.

Practical implications – Some organisations have cultures of predictability and control; others have cultures that recognise their external environment as fundamentally unpredictable, and hence there is a need for responsive, decentralised PMM. Some have sought to change their culture and PMM. Being attentive to how cultural logic affects decision making can help reduce the misalignment in PMM.

Originality/value – A novel contribution is made by applying decision theory to PMM, extending the theoretical depth of the subject.

Keywords Performance measurement, Sustainability, Supply chain management, Case study, Alignment, Decision processes

Paper type Research paper

Introduction: new theoretical perspectives on performance measurement and management (PMM)

The inevitability of change is a fundamental aspect of the modern business environment. Now, more than ever, such changes seem more significant and no less unpredictable. A volcanic eruption disrupts flight schedules, a stock market crashes, socio-political events like the UK’s Brexit referendum, terrorist atrocities or even the rapid impact of disruptive technology on previously stable and mature markets are all issues that can require action by senior managers. Some of these issues may impact on the strategic direction of a business; others may just require flexible response, providing the organisation is sufficiently resilient.

How well an organisation conducts its PMM in relation to the external environment and strategic direction can be described in terms of the PMM system’s “alignment” or “fit”...
Kolehmainen, 2010; Melnyk et al., 2010). If the external environment is stable and changes only slowly, this fit may be easier to establish than if the external environment is turbulent and unstable. Given these two different types of external context, how can PMM remain responsive to change both from outside the organisation and internally, as adjustments are made to the organisational strategy (Micheli and Manzoni, 2010)? Do organisations reformulate their PMM system in response to external changes? In a major review of PMM, Bititci et al. (2012) identify that the context in which PMM occurs is indeed changing and that commonly accepted PMM practices showed limited effectiveness due to the lack of a holistic approach that addresses potential external complexity.

Melnyk et al. (2014) follow this with a Delphi study including high-level practitioner input that highlighted manager’s paradoxes in dealing with reformulation of business strategy and incorporating those adjustments in their PMM system when operating in a turbulent and fast-moving business environment. The research highlighted the need for a conditional, contingent response, given that this is not discussed in the current PMM literature. The study also reported on the lack of fit between business strategy, organisation culture and static PMM systems when faced with a complex and unpredictable external environment.

These two reviews of PMM contain common themes around the role of, on the one hand, complexity (defined by Nicolis and Prigogine (1977) as the unplanned creation of order from turbulence, also referred to as autopoiesis – meaning spontaneously self-organising), which is a form of unpredictability, and, on the other hand, simple, mechanistic, prescriptive guidance, centred on prediction and control. The latter is a core theme of management science since the work of Taylor (1911) and which was extended by Forrester (1958). Following calls for stronger theoretical foundations for PMM research (Richard et al., 2009), this paper seeks to contribute insights into the former question of complexity and unpredictability affecting PMM by drawing on the literature within the field of decision theory (DT). PMM may be described as a form of decision support system, synonymous to a body’s nervous system, that organisations put in place to connect with business strategy, improve effectiveness and achieve desired outputs (Melnyk et al., 2014). The field of PMM may thus find deeper conceptual foundations by considering literature from the field of decision theory, especially that encompassing the topics of complexity, autopoiesis and resulting unpredictability.

**PMM and decision theory in an age of complexity**

One of the seminal works of decision theory is Simon (1947), “Administrative Behaviour: A study of decision-making processes in administrative organization”. This broke with the prior scientific management approach of Taylor (1911) by showing that managers could not perfectly control organisations because they lacked perfect knowledge. In recent years, decision theorists (French, 2012; French et al., 2009) have found the Cynefin framework (Kurtz and Snowden, 2003; Snowden, 2000; Snowden and Boone, 2007) to be a useful summary of the role of simplicity and complexity underlying organisational decision making. It is a sense-making framework that provides a typology of knowledge to assist managers in interpreting the nature of the context of various decision problems. The nature of the framework is shown in Figure 1 and Table I. Four domains are described where the level of knowledge about an external context is outlined, along with recommendations as to the appropriate way to respond.

Whilst applied in various fields from information technology to sociology (French, 2012; Herington et al., 2015; Pelrine, 2011), no consideration of the Cynefin framework has been found in the PMM or wider operations management literature. However, a direct similarity is found with the performance alignment matrix of Melnyk et al. (2014), a guiding framework for managers to consider strategy alongside PMM when operating in turbulent and volatile external environments. To resolve the problems faced by managers in reformulating strategy and PMM, the performance alignment matrix addresses how organisations relate to their external environment and how well the internal PMM aligns with the strategic goals of the
organisation, plus how well they respond to the external business environment. Further definition and exposition is provided in the background literature section below.

The performance alignment matrix is a framework regarding the use of measurement that acknowledges the nature of complexity in organisations “that have multiple levels of decision-making and diverse operating contexts” (Melnyk et al., 2014, p. 182). Furthermore, Melnyk et al. (2014) describe one of the axes of the performance alignment matrix in
terms of certainty. Certainty means being able to give specific outcomes, but under conditions of turbulence or fluctuations in the external environment, uncertainty prompts a flexible approach, stating objectives in general terms, not specific ones.

Decision theory, and in particular the succinct summary of sense-making in diverse contexts provided by the Cynefin framework, addresses the influence of organisational hierarchy and operating context on decision making (Figure 2). The types of sense-making provided in the Cynefin framework are also distinguished on the basis of the certainty of the decision maker; what is certain or known (and thus formalised into standard or instinctive operational responses), through what is knowable (after investigation via classical scientific method or expert analysis), through to what is only retrospectively knowable, due to complexity, or unknowable, due to chaos, reflects a similar spectrum.

Furthermore, both the performance alignment matrix and Cynefin are concerned with the practical issues of decision making in real-world applications. The similarities between the two are thus readily apparent, and correspond closely (see Table I). Given that the academic origins of each are different, there is a potential for deepening the theoretical foundations of PMM via an attempted synthesis of the two, providing a novel contribution to the PMM field through elaborating the existing theory.

The research question is thus stated as:

**RQ1.** How do decision theory concepts such as the Cynefin framework help elaborate theory in PMM in order to better respond to challenges of unpredictability and complexity and subsequent problems of misalignment between organisational processes and the external environment?

The contributions of this paper are threefold. First, we bring the decision theory concepts to PMM, and from our literature searches it seems we are amongst the first to do so. Second, we scrutinise such concepts through an empirical study entailing multiple case studies, and find supporting evidence for the inclusion of decision theory in PMM. Finally, we illustrate the cases using the Cynefin framework, and show how this might assist practitioners in understanding and responding to unpredictable environments when they measure and manage performance in their organisations. The role of organisational culture, defined in terms of dominant logic (DL), is central to shaping this response.

![Figure 2. Hierarchy of decisions and Cynefin framework. Adapted from French et al. (2009) to show additional characteristics](image-url)
The structure of the paper is first to outline the various concepts being considered and provide clear definitions. A conceptual framework diagram is provided (Figure 3) that links the various elements. The research methodology and accompanying research design to explore the applications of PMM alignment using decision theory is then described. Findings and data analysis are then described, followed by a discussion of the findings and implications for future research and practical application.

**Conceptual foundations**

*Challenges in existing PMM systems*

Harkness and Bourne (2015) discuss complexity as a barrier to PMM in terms of two aspects, the first of which is the complexity of the external environment. This is described as dynamic not static and as an "open system" not a "closed system". Second, measuring performance is seen as being hampered by the factors including ambiguity, lack of control, unpredictability and unintended consequences. Acknowledging the presence of these as a challenge to PMM prompts deeper research into the origins of these concepts and hence an interdisciplinary bridge to equivalent discussions in the adjacent areas of study. This is referred to by Denyer et al. (2008) as "reciprocal synthesis".

By incorporating decision theory into PMM, a deeper understanding is possible. A relatively recent summary of the decision theory field and its historical development is provided by French et al. (2009). The barriers to effective PMM noted by Melnyk et al. (2014) echo the definition of bounded rationality established by one of the seminal figures in decision theory – Herbert Simon. Bounded rationality is an essential part of the theory of business administration, complementing and contrasting with neo-classical economic theory of rational decision making (e.g. Taylor, 1911), because in practice, decision makers in organisations find, “their decision making processes are molded by limits on their knowledge and computational capabilities” (Simon, 1947, p. 20).

For PMM, bounded rationality provides a constraint on performance measurement in terms of the accuracy, availability or relevance of data. Performance management may then be affected by bias, where the decision is affected by cognitive factors on the perception of data and what it means. One further effect, studied by agency theory (Ross, 1973), is that of gaming the metrics in...
order to manipulate rewards offered by management (Moynihan, 2008). In other words, manipulating the performance through metrics adjustments in a way that triggers rewards or prevents the triggering of punitive responses by management. The actions of employees can thus become motivated by working to the PMM and the structures it creates, rather than to the traditional goals or objectives of the organisation. This may result in creation of wrong behaviour and mind-sets in an organisation where employees over-perform or under-perform to meet the targets against the assigned metrics set by management. Tett (2009) provides an account of this phenomenon in major banks during the lead up to the 2008 financial crisis. Such issues are common topics in the study of behavioural decision making (French et al., 2009).

The problem of too much focus on metrics in the approach to PMM (Bititci et al., 2012) are analogous to those discussed in the literature on systems theory (Forrester, 1958; Kumar and Kumar, 2016), which is not referenced in the PMM literature, but that have become explicit in the decision theory literature (discussed below) and the operations and supply chain management literature. Notably, it is the divide between general systems theory, dating from Forrester in the 1940s, and that of complex adaptive systems (Choi et al., 2001; Nair et al., 2009; Pathak et al., 2007), which is illustrated in the Cynefin framework. Rather than saying that everything is complex, Cynefin says that sometimes things are complex and sometimes they are not, and it is managerial sense-making of this context that is important.

One of the important question raised by PMM researchers (Bititci et al., 2012; Kennerley and Neely, 2003; Kennerley et al., 2003) was how to develop and maintain a dynamic PMM system over time in the climate of market volatility and turbulence. Kennerley et al. (2003) proposed a PMM maturity model that encapsulates the role of external factors such as market volatility and regulations, and internal factors such as culture, people and process in effectively managing and developing company’s PMM system that truly represents the organisation’s changing requirement. The paper addresses the “what” questions linked to dynamic PMM but had limited discussion on “how” to adapt and operationalise the new measures in a highly turbulent environment. However, this work could be considered as one of the early papers to debate the need for dynamic PMM.

Kolehmainen (2010) further picks up this theme to draw insights from a large telecommunication company and comments on the role of individual managers in adapting the PMM in a dynamic market environment. Flexible and strategically aligned PMM will allow different parts of the system to work together effectively and empower individual managers to adapt to turbulent market conditions. Such a system moves beyond command and control measures to measures that promote organisational learning (Kolehmainen, 2010; Micheli and Manzoni, 2010).

Pekkola et al. (2016) address the need for a dynamic PMM system called for in the literature (Bititci et al., 2012; Garengo et al., 2005; Kennerley and Neely, 2003; Melnyk et al., 2014; Nudurupati et al., 2011) by proposing a framework for flexible performance measurement system design for small to medium enterprises operating in a turbulent environment. Here, the core permanent measures of the framework are cost/profit centric and support measures aligned with the strategic goals. The supportive measures need to be flexible and dynamic to change as the organisation changes its strategic goals and vision when operating in the turbulent market conditions. However, this framework is too simplistic and lacks integration with decision theory, although it is a good starting point for small to medium enterprises to develop a flexible and robust PMM system.

As discussed by Melnyk et al. (2004, p. 212), the failure to respond to a dynamic environment causes stresses between the organisation, the external environment and the organisational strategy, which can become a source of conflicts between a firm and its customers or suppliers. The direct fit between environment and PMM has received less attention in the literature. Melnyk et al. (2014) address this by defining the “performance alignment matrix”. This is a newly created framework that categorises specific and general
outcomes and specific and general solutions, which relate to the degree of certainty that defines management setting of outcomes (i.e. specific vs general outcomes) and specific approaches the organisation adopts to deliver the outcome (i.e. solutions).

Melnyk et al. (2014) classifies the performance alignment matrix as shown in Table I, above: measurement-driven management (specific outcome and solution), outcome-driven solutions (specific outcome and general solution), assessment-driven management (general outcome and solution) and solution-driven outcomes (general outcome and specific solution). The performance alignment matrix highlights that a failure to achieve alignment, or “fit”, between the internal PMM system and the external environment is a major barrier to effective PMM. Whilst providing a strong response to issues of predictability and unpredictability in PMM, such ideas are also present in decision theory fields and in particular the Cynefin framework, which is discussed in the next sub-section.

**Decision theory and Cynefin framework**

As described below, decision theory is one relevant area of business and management studies to examine complexity, uncertainty and unpredictability, alongside strategic management (Mintzberg and Waters, 1985; Mintzberg and Westley, 2001). In their account of the different branches and the evolution of decision theory, French et al. (2009) highlight the Cynefin framework (Kurtz and Snowden, 2003; Snowden, 2000, 2002; Snowden and Boone, 2007) as a useful and effective summary model (see Figure 1). Cynefin is a Welsh word (pronounced ku-ne-vin) roughly meaning habitat, including social and cultural elements as well as environment. It “signifies the multiple factors in our environment and our experience that influence us in ways we can never understand” (Snowden and Boone, 2007, p. 71).

The framework thus concerns the subtle perception of a decision context and is a typology of knowledge in relation to decisions, based on the nature of their context (French, 2012; French et al., 2009). It acknowledges the difference between the structured and predictable decision contexts, where managers can exert traditional command-and-control, and the unstructured and unpredictable decision contexts where they must either allow order to emerge from a group-oriented level, or attempt to impose order on apparent chaos. Figure 2, from French et al. (2009), links the epistemology of decision context shown in the Cynefin framework with the decision authority in a conventional organisational hierarchy (shown on the left side). PMM can sit at the level of tactical and operational decisions, ideally aligned with the higher level of strategic decision making.

The two sides of Cynefin (Figure 1) reflect a divide in epistemology (the philosophy of knowledge); the divide between rationality and bounded rationality defined by Simon (1947) in relation to management decision making, reflects this, and deliberately contrasts with Taylor (1911). As shown in Figures 1 and 2, the Cynefin framework consists of four domains, grouped into two halves of structured and unstructured domains. When a decision context is sufficiently structured and stable, it enables quantitative analysis. When it is characterised by dynamic change, plural and contested definitions, or non-linear complexity (Lorenz, 1963; Prigogine and Stengers, 1984), the space is unstructured (Kurtz and Snowden, 2003; Snowden, 2000). This offers a means to determine where it is possible to make rational decisions or where bounded rationality prevents this, and thus alternative methods are needed. The structured is further divided into simple and complicated domains, and the unstructured is divided into complex and chaotic domains.

Each domain suggests responses, detailed in Snowden and Boone (2007), that managers should take when faced with the different decision spaces defined by each domain. This responds to the concept of fit, concerning how well-aligned internal processes are to the external environment. The lines between each domain are curved to emphasise that the borders between the domains are contextual, and that the figure is not a graph with two axes marking quantitative variables. The Cynefin framework is effectively a typology of
knowledge management, called a social ecology of knowledge (Snowden, 2000), in order to consider different approaches to organisational knowledge.

The four domains are also intended to illustrate that misalignment can occur when the external context of a decision may shift from one domain to another in a potentially dynamic fashion. The shift from a simple-structured domain to an unstructured-chaotic domain is for instance illustrated by a line symbolising a breaking wave on Figure 1. A simple bureaucratic system may cease to function effectively if the underlying assumptions shift and the context becomes increasingly structured or that the structure starts to break down into turbulence. It may become gradually more complicated, as additional requirements are added to a project for instance. Increasing the numbers of stakeholders involved in a project can result in a breakdown of shared perception, and this divide can result in an otherwise stable and structured context becoming less structured. These correspond to the concept of fit and subsequent description of the performance alignment matrix in Melnyk et al. (2014), which states that how a PMM responds to change is affected by organisational culture. PMM is thus both “a technological process and a social one [...]” (Melnyk et al., 2014, p. 3).

Managerial sense-making and response is central to Cynefin and echoes the strategic management concept of dominant logic (DL) (Bettis and Prahalad, 1995; Prahalad and Bettis, 1986). This is described as the dominant mindset or view of the world possessed by top managers that shapes their decision making, and through their shaping of organisational culture, affects the decision making of others in the firm. DL is a schema based on the beliefs established as a result of employees’ experience within given firms and industries. Prahalad and Bettis were concerned with the role that DL plays when a firm in one industry acquires a firm in another industry and misunderstandings result, leading to underperformance. The perceptual, cultural and sense-making aspects of DL relate to how decisions are made in practice. This is important to understanding the three aspects of strategy, PMM and organisational culture mentioned by Melnyk et al. (2014) in terms of fit. In this paper, the Cynefin domains are taken as descriptions of DL, providing a decision theory perspective on organisational culture, which can then be applied to the issue of PMM and fit. Figure 3 expresses a conceptual framework covering the relationship between these key concepts.

Awareness of an organisation’s performance, captured via data, enables control, and as such the ability of managers to direct organisations towards meeting strategic requirements. Yet, if the context is unstructured as a result of unpredictability, then this approach is problematic. Cynefin suggests that under unstructured contexts, an approach based on responsiveness instead of forecasting, or emergence instead of control is required. PMM should reflect this dynamic, but the issue appears under-theorised in the literature, as noted by Melnyk et al. (2014) or Harkness and Bourne (2015), which address the role of complexity as a problem for PMM, but less so on how to respond to it based on the existing theory in this space.

*Mapping the performance alignment matrix with the Cynefin framework*

The performance alignment matrix framework is found to have a useful precursor in Snowden’s Cynefin framework. Introducing this framework and its associated conceptual underpinning helps provide a greater theoretical depth to the topic of PMM. This answers recent calls for deeper understanding of the role of volatility, uncertainty, complexity and ambiguity (Bennett and Lemoine, 2014) and how they are addressed as aspects of the external business environment, interpreted (via sense-making) in different ways by the internal organisational culture.

As stated by Melnyk et al. (2014), “for PMM to be effective it has to fit the environment in which it operates” (p. 183). Similarly, the Cynefin framework is built on an empirical observation; “Wise executives tailor their approach to the complexity of the circumstances they face” (Snowden and Boone, 2007). A range of sources of unpredictability, including volatility, uncertainty, complexity and ambiguity, have long been recognised as characteristic of manager’s perception of the external environment (Bennett and Lemoine, 2014; Simon, 1947).
However, the Cynefin framework offers valuable insights in showing that such unpredictability is not always present. Many circumstances are readily predictable, and hence the decision makers’ “approach” should be tailored to the circumstances of the external environment in order to be effective – there is no “one-size fits-all” approach.

The failure of sense-making (Browning and Boudès, 2005; Weick, 1995) in relation to changing circumstances described by Cynefin is again similar to what Melnyk et al. (2014) call a “lack of fit” or misalignment. Notably, in defining the performance alignment matrix as a framework to describe alignment between PMM and the external environment, the authors cite precursors in strategic management; Venkatraman and Camillus (1984) describe strategy in relation to organisation theory’s open systems view where, “strategy [can] be conceptualised as a pattern or stream of decisions taken to achieve the most favourable match or alignment between the external environment and the organisation’s structure and process” (Venkatraman and Camillus, 1984, p. 516). This alignment, or fit, is thus a phenomenon that has been conceptualised over time in the management literature, particularly in strategy (Chorn, 1991; Henderson and Venkatraman, 1993; Walter et al., 2013). Both the Cynefin framework and the performance alignment matrix seek to conceptualise the phenomenon but do so from slightly different backgrounds: strategic management and organisational performance in the case of PMM, and knowledge management and decision making in the case of Cynefin.

The PMM concept of fit, as illustrated in the performance alignment matrix in Melnyk et al. (2014), can be elaborated using the more detailed explanations and managerial responses described in the Cynefin framework. Such decision theory concepts are empirically examined through case research into the implementation of corporate social and environmental responsibility practices via sustainable operations and supply chain management (SOSCM) and related PMM. Seven firms in different sectors are investigated to find how DL and the external environment are aligned and the subsequent alignment of PMM around SOSCM. The findings lend support to the adoption of the Cynefin framework as a useful means to understand how organisations should perceive and respond to the external environment. The next section provides more information about the method adopted for this research.

**Research method: theory elaboration and synthesis**

The initial approach of combining the performance alignment matrix with Cynefin is based on the work of Denyer et al. (2008) which describes various approaches to theoretical synthesis such as reciprocal synthesis (different descriptions of the same thing) or lines-of-argument synthesis (different descriptions of slightly different but connected things). However, as there has to date been no application of Cynefin within PMM or operations management, the research approach taken here was to consider theory alongside original empirical data collection using the methodology of Ketokivi and Choi (2014). Here case studies are used in tandem with a known middle-range theory (Soltani et al., 2014) in order to explore “theory elaboration” (as contrasted with novel theory generation or large-n statistical theory testing). Multiple cases across different industries are used to consider how PMM is undertaken in different external contexts. Focussing on a specific area of PMM, firms’ SOSCM performance practices are taken (as opposed to, say, PMM in human resources management). Bititci et al. (2012) note the rising importance of sustainability factors in PMM, and various papers cite problems of bounded rationality, such as plural and contested definitions in SOSCM (Hahn et al., 2014; Preuss and Walker, 2011), or the temporal and physical distance between cause and effect often found in sustainability issues (French and Geldermann, 2005). A review of how such decision theory concepts are addressed in sustainable supply chain management research is provided by Alexander et al. (2014).
As discussed by Eisenhardt (1989), such case research uses theoretical sampling rather than statistical/random sampling. Cases are selected because they represent high variance in the relevant concepts, not high levels of comparability as with the matched-pairs approach (Yin, 1994). Novel ideas emerging from the empirical data enable theory to be modified to reflect these findings. Miles and Huberman (1994) define this as the balance between a tight design and a loose design. In the former, concepts are pre-specified, but in the latter, they are not well-defined in advance, enabling emergent ideas. As such, each case company was selected via a process of polar sampling of key variables, including those established via iteration and juxtaposition (Eisenhardt, 1989; Yin, 1994). This establishes a range of different internal and external characteristics from which to consider levels of fit. Triangulation as a means of verification and validation was conducted via the use of multiple data sources per case and across cases (Pauwels and Matthyssens, 2004).

The DL in the organisational culture and the external context are established through the data collected using critical realist abstraction from data to underlying concepts and causal mechanisms (Rotaru et al., 2014). Dubois and Gibbert (2010) provide additional justification for this approach by highlighting the importance of alignment between theory and method. Given that the Cynefin framework concerns bounded rationality, and that the theoretical model of applying this to PMM is nascent, then a quantitative methodology would not be appropriate at this stage. Furthermore, as the conceptual model is nascent, exploratory work founded in empirical field research is preferred over theory testing, or pure conceptual work derived in the absence of practical application.

Empirical data were collected over a two-year period from seven UK-based international companies and their supply chains across different sectors (Table II) contacted via a variety of professional networks. All had adopted SOSCM policies and practices and all interviewees were centrally involved in them. The point of saturation (Eisenhardt, 1989) was seen as met in the seventh case, so further potential cases were not pursued. Data collection used the semi-structured elite interview technique (Vaughan, 2013), whereby researchers can interrogate deep into the topic whilst also maintaining flexibility to encourage new topics to emerge during discussion. In addition, meetings were observed to gain deeper insight into how particular processes influenced decision making, including meetings between companies and their customers and suppliers, and between different departments and subsidiaries in a firm. All interviews were conducted under conditions of anonymity, enabling candid responses but requiring that information allowing firms to be identified has had to be kept from publication.

Examples of questions asked via semi-structured elite interview technique:

- Introductory questions: describe the company and where it is at presently. Describe your projects in relation to sustainability – both internally and in the supply chain.
- SOSCM and PMM questions: how do you measure performance in relation to sustainability – both in internal operations and in the supply chain?
- Decision theory questions: what processes for decision making are there in the organisation? Can managers act on their own intuition or do they need to follow clear rules and provide particular evidence before being able to act? Are PMM and SOSCM issues simple and easy to understand/involves lots of variables and expert analysis/are subject to unpredictability? Are there aspects of PMM and SOSCM that cause problems?

Around 52 hours of interview data were recorded, transcribed and coded using descriptive and conceptual coding (Saldaña, 2012). Recursive abstraction is used to derive codes from the in vivo data with consideration given to processes of establishing causal foundations from interview data in Rotaru et al. (2014). Coding was conducted via repeated re-reading of
the interview transcripts to isolate the relevant concepts and variables. Two researchers independently coded transcripts and compared results to ensure accurate coverage and reach consensus to establish validity in the coding protocol.

The process for reliability and validity determined by Pauwels and Matthyssens (2004) was used, as outlined in Table III. This updates and extends the validity processes provided by Yin (1994), Miles and Huberman (1994) and Eisenhardt (1989).

**Findings**

Table IV shows the DL coding of each organisation, using the Cynefin domains from Figure 1. These are also plotted on Figure 4. Illustrative quotes from the data are also included in Table IV with references to sample quotes provided in brackets, cross-referencing data provided in Table AI. Having determined the DL for each firm, Table V then illustrates the alignment with the external environment under typical conditions, and then in relation to SOSCM issues, including instances of misalignment.

As noted in the strategic management literature cited above, alignment or fit is a concept about how the internal logic and process perceive the external environment. As a sense-making
framework, Cynefin, by definition, concerns the link between the internal and the external. A short narrative description of each case follows describing the DL using Cynefin domains (expressed in the text and Tables IV and V as Domain 1, Domain 2, etc.). The rich description includes the way in which SOSCM was incorporated into PMM. The next section then discusses the theory elaboration for PMM derived from the case data.

<table>
<thead>
<tr>
<th>Criteria for reliability and validity</th>
<th>Description and actions taken to meet the criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical sampling</td>
<td>Selecting informants on the basis of theoretical relevance</td>
</tr>
<tr>
<td></td>
<td>Iterative polar sampling of relevant variables (sector, size, dominant logic for decision making, external context) (Eisenhardt, 1989)</td>
</tr>
<tr>
<td>Triangulation</td>
<td>Reduce potential for bias or error by:</td>
</tr>
<tr>
<td></td>
<td>Elite interview technique (Vaughan, 2013)</td>
</tr>
<tr>
<td></td>
<td>Synchronic primary data source triangulation (different respondents, same topic).</td>
</tr>
<tr>
<td></td>
<td>Diachronic primary data source triangulation (same respondent, multiple occasions).</td>
</tr>
<tr>
<td></td>
<td>Comparison of primary and secondary data</td>
</tr>
<tr>
<td></td>
<td>(Pauwels and Matthyssens, 2004)</td>
</tr>
<tr>
<td>Pattern matching</td>
<td>Multiple, independent coding of transcripts</td>
</tr>
<tr>
<td>Analytical generalisation</td>
<td>Comparison to existing theory in tandem with data collection via progressive focussing (Ketokivi and Choi, 2014; Sinkovics and Alfoldi, 2012)</td>
</tr>
<tr>
<td>Validation</td>
<td>By juxtaposition and iteration (cross-case comparison) (Pauwels and Matthyssens, 2004)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>Description of organisational culture and strategy</th>
<th>Dominant Logic</th>
<th>Quote ref. (Table AI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Organisational culture shaped by majority of employees being electronic engineers. Rules-based decision making. Work is focussed on precise technical specifications. Low tolerance of ambiguity or uncertainty. Core strategy is high quality (not low cost)</td>
<td>Domain 2</td>
<td>Q1.1</td>
</tr>
<tr>
<td>2.</td>
<td>Organisational culture is shaped by strong ethical values. Principles-based decision making. Core firm strategy is serial innovation strategy based on driving improved ecological and ethical characteristics of products to align with customer values (this can be described as a quality strategy with eco and ethical elements regarded by customers as quality characteristics)</td>
<td>Domain 3</td>
<td>Q2.1</td>
</tr>
<tr>
<td>3.</td>
<td>Organisational culture centred around decentralised decision making and tolerance of uncertainty. Analytics not part of the organisational culture</td>
<td>Domain 3</td>
<td>Q3.1</td>
</tr>
<tr>
<td>4.</td>
<td>High level of rules-based analytic decision making, but introducing a decentralised decision-making approach</td>
<td>Domain 2, but introducing Domain 3</td>
<td>Q4.1</td>
</tr>
<tr>
<td>5.</td>
<td>Organisational culture shaped by high context specificity and variability of construction work, leading to high level of danger, requiring decentralised decision making responsibility</td>
<td>Domain 3</td>
<td>Q5.1</td>
</tr>
<tr>
<td>6.</td>
<td>Organisational culture typical of large engineering-based firm defined by regulated standards and customer-base similarly subject to regulation</td>
<td>Domain 2</td>
<td>Q6.1</td>
</tr>
<tr>
<td>7.</td>
<td>Core business of the firm is a low variation chemical engineering process requiring expert knowledge and operating in highly regulated, controlled conditions</td>
<td>Domain 2</td>
<td>Q7.1</td>
</tr>
</tbody>
</table>
**Case 1: electronics firm**

Case 1 is a leading designer of electronic components for popular consumer devices. The large majority of the firm’s employees were engineers. The external environment was relatively stable and simple, with incremental improvements in product performance delivered year on year with little radical innovation. The DL was clearly structured and rules-based, shaped by scientific processes, so corresponding to Cynefin Domain 2.
SOSCM issues had become highlighted in the sector, with recent regulations imposed via the US Dodd-Frank Act. SOSCM issues had thus been introduced as a customer requirement. However, given the nature of the firm and its position in the supply chain, this was outside of the firm’s expertise or ability to influence (Q1.2). It was not an element included in the technical specifications that they passed up the chain to their suppliers in manufacturing and assembly. Hence, the response to SOSCM at the time of the interviews was bureaucratic and procedural (Domain 1), focussed on the legal liability of disclosure, requiring suppliers to confirm compliance, but having no mechanism of auditing or enforcement. The regulation resulted in minor changes in specification that reduced the potential liability but did not engage in supplier capacity building, or extending their engineering expertise to consider SOSCM issues more fully (Q1.1-Q1.5). Under the performance alignment matrix (Table I), this approach represented “measurement-driven management (specific outcome and solution)”. 

Case 2: FMCG manufacturer
By contrast to Case 1, Case 2 had an SOSCM policy fundamental to the founding of the firm some 20 + years ago. The firm is an international manufacturer of ethical and ecological consumer goods. The interviews found a very strong organisational culture based around principles designed to maximise SOSCM performance in order to align with pro-ethical/pro-environmental customer values (hence, the strategic position of the firm). Notably, while Domain 2 processes existed in the lab and factory floor, the firm claimed insufficient resource for Domain 2 analytics in other aspects of management, including marketing and forecasting (Q2.3). Instead, the DL was clearly defined by Domain 3 logic (Q2.1, Q2.2). Characteristics of this were a decentralised approach to decision making inherent in PMM processes. A wide range of employees could be called in to work on decisions outside of their job roles via a principles-based approach to decision making (Q2.1, Q2.2, Q2.3). Firm strategy was essentially to align its performance with the ethical and ecological values of their consumers, establishing a form of quality strategy where exemplary SOSCM performance was fundamental to competitive advantage. Under the performance alignment matrix, this approach compared to the “assessment-driven management (generic outcome and solution) domain where employees are given freedom to take decisions without too many constraints. However, while the DL was well-aligned to PMM, the firm stated that it lacked the resources to conduct Domain 2 analytics as some of its larger competitors did. Despite a niche position, because the sector of fast-moving consumer goods is mature and essentially stable, the Domain 3 position may be aligned to strategy, but misaligned to the external environment given the superior performance possible from a Domain 2 capability aligned with a stable competitive environment.

Case 3: restaurant chain
Case 3 is a large international restaurant chain with a strong, values-focussed culture intended to create the atmosphere of a family-owned restaurant despite a large size (Q3.1-Q3.2). The DL is Domain 3 as decentralised decision making and values of personal responsibility are central. With a strategic position of low cost but requirement for fresh produce (reducing the potential for inventory to smooth out variability in customer numbers) purchasing decisions were decentralised to the branch level. Local, granular knowledge and the instincts of branch managers were assumed to be required; knowing that a local sports team was playing on a particular day would prompt expectations of a busy day, rather than centralised ordering on the basis of forecasting via extensive Domain 2 data analytics.

Like Case 2, this firm lacked the Domain 2 analytic capability of a dedicated operational research department, and although it also had a principles-based, decentralised culture for
decision making, it had this for very different reasons to Case 2, based more on the nature of the external environment as being characterised by high variability, than on resource constraint and strategic need for customer values-alignment (beyond the atmosphere of restaurants).

Two issues of SOSCM were explored; the first was a plan to reduce vulnerability to supply chain disruption. This was prompted by the EU horsemeat contamination scandal of 2015 but then applied to vulnerability to future climate change impacts. The inherent unpredictability of future disruptions did not affect investment in such initiatives because the Domain 3 DL of the firm had a high tolerance of ambiguity as to the business case. Such initiatives were supported by the board on the basis that if it sounded like the right thing to do, then it should go ahead (Q3.2).

The second SOSCM issue was the introduction of government carbon taxes (via the Energy Savings Opportunity Scheme). This was addressed as a calculation from the accounts department where total spending on electricity and fuel was converted into the associated carbon footprint and related government levy. This is a low-level, Domain 2 analytic process, or advanced, Domain 1 bureaucratic process. The policy did not however prompt a reduction in organisational carbon pollution as perhaps intended as to do so would require the firm to have a detailed data architecture, recording the operational consumption of every different location of the business. The size of the levy did not warrant the cost of developing a PMM system to target cost-effective reductions in energy consumption. This would be a more sophisticated Domain 2 undertaking, requiring expert analysis of a large number of variables. Instead, the carbon tax was seen as a cost of doing business (Q3.3). Latter firms in this study maintained a similar position towards this scheme. Misalignment can thus be considered in the SOSCM PMM as this is imposed by the government, as a broadly Domain 1 bureaucratic undertaking with some expert analytic Domain 2 capability needed, whereas achieving improvement in carbon performance would require an extensive and costly Domain 2 analytic approach involving the creation and study of potentially large sets of data to determine optimum performance output balanced against particular types of energy efficiency intervention (Q3.3).

While both Cases 2 and 3 had conventional, cost-based performance measures, it is notable that both were tolerant of ambiguity in decision making. This relates to the definition of a Domain 3 DL regarding level of certainty, providing some empirical validation of the Cynefin framework.

Case 4: high street finance firm
Analytic sensitivity to profitability was seen in Case 4, a stock market-listed international finance firm. The firm exhibits a Domain 2 DL as decisions were typically made on the basis of a complicated series of internal policies and regulatory standards, requiring bureaucratic responses and levels of expert analysis (Q4.1). PMM around SOSCM were well implemented and had been for many years, showing alignment between this aspect of PMM and the DL. Strong alignment between PMM for SOSCM and some strategic objectives could be seen in, for instance, energy efficiency programmes helping to improve productivity and profitability by reducing waste of energy. Similarly, local sourcing helped with local economic development, which improved reputation.

One other major SOSCM initiative was impacted by unpredictable volatility. A clear belief by the firm and its suppliers of linear energy price inflation (Q4.2) was affected by a sudden, unexpected drop in the price of oil a few weeks after one interview. A purely economic rationale for carbon reduction expressed in the interview was then undermined by the impact of this energy price fall on the anticipated payback periods for investment in energy efficiency measures. This represents a Domain 2 analysis becoming misaligned to an unpredictably changing external environment. Such issues of price volatility are perhaps
inevitable factors of business, so alternative sense-making may have been cautious of such inevitability. In Case 3, the DL tolerated ambiguity about future paybacks, reinforced by the principles-based decision making. The key issue is not only how swiftly the firm responds, including how the PMM responds, but how the firm might anticipate such fluctuations and prepare accordingly.

A second key finding from Case 4 relates to a misalignment between customer perception of the business, the firm’s strategic goals around profitability, and the implications for DL. The firm was selected on the basis of its large size, and sophistication of its PMM systems, as a polar contrast to the small size of the previous three firms (as required by principles of theoretical sampling for high variation). However, a serendipitous discovery raised unprompted during the interviews was that the firm sought increased responsiveness to customer needs by introducing an alternative DL (Q4.1). This aimed at shifting from a Domain 2 structured, rules-based approach to a decentralised, principles-based Domain 3 logic.

Targets imposed by the former PMM were considered responsible for rising levels of customer dissatisfaction, which undermined profit targets and associated long-term strategic goals. Loosening the PMM system by shifting to a Domain 3 logic meant allowing customer-facing staff more discretion over decisions on the basis of their intuition about possible reputational impacts, rather than the pressure to hit certain department level performance targets. At the time of the data collection, this change management programme in the DL was at an early stage. Future research could examine the success of this, given the potential challenges of changing an organisation’s DL.

Case 5: construction contractor
Case 5 is a major contractor in the construction sector who showed a very strong Domain 3 DL, particularly in relation to project management and health and safety. Because construction sites are constantly changing on a daily basis and schedules and workflow are impacted by unpredictable issues like the weather, in ways that factory or office work is not, they require that project managers are able to respond to frequent change (Q5.1, Q5.3). They are also very dangerous places to work; from 2001-2011, 760 workers were killed on UK construction sites, compared with 448 British Army soldiers killed fighting overseas over the same period (HSE, 2016) (Q5.2). As such, a very strongly decentralised decision making culture had developed where people follow common sense principles instead of (blindly) following prescriptive rules.

The strength of this Domain 3 DL resulted in an unprompted account of the dangers of the balanced scorecard PMM system by major clients. Here, the misalignment between PMM and strategy resulted in serious problems for clients’ real estate portfolios; purchasing of land (deemed a performance metric) continued when macro-scale demand declined very rapidly after the 2008 financial crash. The misalignment resulting from this volatility has had a lasting impact on the real estate sector and the performance of the firm and its competitors in the sector (Q5.5).

In terms of SOSCM, the misalignment with PMM systems was also seen with the adoption of a government-promoted eco-standard for building design. The Building Research Establishment Environmental Assessment Method (BREEAM) system had become a widespread extension of updated mandatory building regulations affecting the environmental performance of buildings. But whilst building regulations were a simple, procedural and scientific measure (requiring expert analysis and so a Domain 2 activity), the BREEAM system consisted of a wide range of potential environmental features assigned scores via a more simple, bureaucratic Domain 1 system. Architects and contractors had to work with this Domain 1 system, even though the act of designing and constructing a building included many Domain 2, and Domain 3 activities. This antagonism was addressed

Complexity in PMM
in various ways, such as an architect firm becoming a certified BREEAM assessor so as to best balance the needs of the bureaucracy with the less mechanical and more aesthetic, qualitative or context-specific aspects of building design (Q5.4).

Theoretical implications of this are that while environmental performance of UK construction has been subject to substantial legislation for more than ten years, misalignment between prescriptive demands and the best-quality outcomes for buildings are clear. Future research could explore the nature of the misalignment in this sector in relation to BREEAM from the perspective of the Cynefin model of PMM described here.

Case 6: heavy manufacturer

Case 6 is a large, international manufacturer of construction products. SOSCM initiatives have become strongly aligned with the organisation’s strategy and PMM systems. As an engineering-based firm, a Domain 2 logic was dominant (Q6.1) and two large projects in SOSCM were investigated. First was the attempt to adopt the widespread use of life cycle analysis (LCA) on products. This involved a process of auditing environmental performance to create environmental product declarations (EPDs), a formal certification scheme. Possessing an EPD was rewarded by a point on customer supplier selection processes (a bureaucratic, Domain 1 activity). Yet, problems were seen in creating Domain 2 analysis of a product’s whole life cycle environmental impacts. With no regulatory or industry-level standard, the processes of LCA are highly variable. Furthermore, manufacturing a similar product out of different materials means highly non-comparable methodologies, assumptions and criteria. LCA is thus found to be a plural and contested Domain 3 undertaking (Q6.2).

Misalignment is thus found in the use of LCA as a PMM. Instead of becoming a Domain 2 system where an accurate and unambiguous analysis between two different suppliers is possible to determine which has the better environmental impact, LCA is in practice in a Domain 3 state. Meanwhile, in practice, because supplier selection processes award a point on the basis of having an EPD but not on any particular performance specified in the EPD, then LCA has become a Domain 1 undertaking. Establishing LCA as a Domain 2 system would require extensive regulatory standards and formal methodologies, independent assessment and auditing processes.

A second example of SOSCM in the PMM of the firm was on initiatives to reduce the carbon footprint of the firm. Two factors in the external environment impacted on this. First, international competition meant a strong downward cost pressure (Q6.3), meaning that energy efficiency measures were encouraged provided that they had a net cost reduction over a short payback time. Therefore, sustainability (implemented via SOSCM practices) as part of a strategic position of quality was problematic. Without such aspects being assisted by mandatory performance measures possible given the globalisation of the market and role of existing technical performance standards (that did not mandate environmental or social performance), the company’s competitive position had to adopt a strategy of low cost.

It was clear from customer tendering processes and weightings given to product specifications against cost, that price was the main decider, not sustainability criteria (or even issues such as product longevity or additional benefits such as service agreements) (Q6.3). This raises many questions for internal and external alignment. The practice of the firm suggested that PMM had become aligned to the new strategy to advance sustainability performance, but this strategy was becoming misaligned with the external competitive environment, and so the PMM for SOSCM refocused around the now better aligned strategy of low cost.

An additional instance of misalignment around SOSCM was found in discussions on the potential for changes to energy supply to low carbon alternatives. The highly unpredictable nature of the UK Government’s policy on this at the time was seen as undermining any potential for justifying investment in certain types of new plant (Q6.4). This contrasts with
the example in Case 4, where there was an assumption of a predictable growth in energy price justifying investment over a given payback period. In Case 6, the larger scale of the energy consumption (given a manufacturing rather than services firm) meant the need for more substantial investment, more sensitive to government energy policies. However, anticipated unpredictability in this policy (rather than anticipated predictability in commodity price) prevented capital investment. Taken as a Domain 4 context, the responses to this unpredictable context in Cynefin are to either wait to see if order reappears over time or to try to impose order (Domain 4 left alone to Domain 3, or imposed to Domain 1). An example of trying to impose order may have been to lobby government to set clear new rules in place (Domain 4 to Domain 1 or 2). In practice, the firm waited without making a decision, delaying progress on this aspect of environmental performance.

**Case 7: chemical processing**

Finally, Case 7 is an engineering firm involved in the large-scale chemical processing of materials. The firm ostensibly operates as a services firm rather than manufacturer as its clients retain ownership of the materials undergoing processing. The firm described its activity as essentially very simple, albeit still in an expert area of chemical engineering and so is determined to have a Domain 2 DL (Q7.1). It only provides a small range of services but does so at a large volume, providing to a substantial share of global markets. SOSCM practices and associated PMM are regulated by government in environmental terms but extended into areas of corporate responsibility via investor requirements such as Global Reporting Initiative standards. As such, SOSCM is well established into key performance indicators (KPI) and embedded into PMM (Q7.2, Q7.3). In contrast to Case 1, this was also a strong engineering firm, and operated according to a clear Domain 2 logic, but it was one that had fully implemented SOSCM into its PMM processes. Two instances of alignment were also found. First, that an accident affecting a major customer impacted global inventory levels, as the production slowed, available supply of materials increased and so demand for their core product was reduced. This unpredictable event meant a change in the external environment, impacting firm strategy by reducing the market price. However, the PMM in relation to SOSCM was not affected, although there were strategic implications for the firm.

Second, when building new facilities, the firm’s stakeholder engagement capability needed to align with the local community concerns. This is a Domain 3 undertaking, and the approach taken of full transparency and open days at existing facilities is designed to improve the general public’s understanding of the business and its operations. This can be theorised as an example of enabling community sense-making so as to move the decision context of the community from Domain 3 plural interpretations (such as whether or not the factories had a risk of pollution or accident) to a common understanding of the high standards for safety and environmental management. This case was deemed the final one as a point of saturation seemed to be reached.

**Discussion: fit between DL, external environment, strategy and PMM**

Many of the cases have a good fit between the DL, the PMM and the external environment. As described in the previous section, some of the case organisations occupy inherently complex environments; specifically, Case 3, the restaurant chain, and Case 5, the construction contractor. These have a DL that explicitly addresses this context, and so represents a strong fit between the DL and the external environment. Similarly, Cases 1 and 7 occupy stable environments and have a DL that is a strong fit with this.

In other cases there are examples of lack of fit, but it is notable that in the Cynefin literature, such as Snowden and Boone (2007), such misalignments result in either...
ineffective operations or attempts to resolve them quickly. Drawing on the work of Mehnyk et al. (2014), where such lack of fit causes a problem, it is because the PMM does not adapt quickly enough to the changing context. Firms that fail to adapt may suffer and so this may demonstrate an instance of bad management. Case 2 had a Domain 3 DL that may be thought of as misaligned with its stable external environment, and hence contributing to a weak competitive position against rivals that did have a Domain 2 DL. However, the Domain 3 DL was thought of as important for aligning with the customer values, which was an important part of the firm’s strategy. Meanwhile, Case 4 showed an example of a Domain 2 DL misaligned with customer values and hence undermining the firm’s strategy. Here, the shift was from a Domain 2 DL to a Domain 3 DL. Alignment with customer values as an aspect of firm strategy may thus be a useful avenue for future research.

A central issue for PMM scholars and practitioners is thus to gain better understanding of the nature of lack of fit between DL, PMM and external environment, and the nature of change. In particular, it is the external environment that is most likely to change, and hence the ability of a firm to reflect this. The Cynefin framework provides useful insight into this process for PMM. Evidence from the data provides an initial empirical validation of the Cynefin framework in this way; however elaboration of the theory is also prompted by the data.

During the course of the investigation, there were clear instances of unpredictable change affecting many of the businesses. The responses to this depend on the underlying assumptions of linearity. Energy efficiency as an instance of PMM (applied in manufacturing production, in offices and retail operation, in vehicle fleets and construction projects) was commonly framed as a driver of cost saving, alongside a common assertion (across cases) that energy prices would only rise. This was an argument that improving environmental performance by investing in means to reduce consumption was going to be competitively advantageous. However, the sudden fall in energy costs (IEA, 2015) undermined the costs of implementing such measures. The examples in Cases 3, 4 and 5 all saw a justification of the cost of implementing these efficiency measures on the basis of an assumed linear increase in the cost of energy. Hence, sense-making around SOSCM and related PMM was clearly that this was a predictable, structured decision problem.

This assumption for a linear trend is common but unfounded. Energy prices have experienced volatility over many decades. A similar example of SOSCM is also seen in Case 6, the construction products manufacturer, where having LCA fails as a process to drive improved ecological performance due to inherent complexity. However, merely having a certificate scores a point as a Domain 1 activity in supplier selection decision processes. Despite trying to generate accurate data, in practice no Domain 2 structured-complicated decision analysis could be developed to establish a clear, rational calculation of the benefit of one product over another, despite this appearing as an assumption among LCA advocates.

While such a contradiction may appear damaging, in this case it is merely the misalignment between an external environment of regulation and the realities of the proposed PMM. In practice, the reaction appeared to be one of resigned indifference. In the case of regulations regarding investments in low carbon electricity generation, the level of unpredictability in the associated policies was so high as to create paralysis in the company, so more a case of exasperation. Both cases highlight that the external economic and political environment is critical and one that firms can either choose to engage in or not. Investors crave predictability, and in its absence investment stalls, highlighting the value of a stable context. The importance of political and macro-economic stability may thus be seen as all important. As discussed by Christopher and Holweg (2017), periods of external stability and linear economic growth can persist for some time, but be punctured by periods of instability and no net growth.
A further important issue to emerge clearly from the data is that ensuring a good fit between strategy and PMM is not in itself necessarily sufficient to maintain long-term survival. A number of cases demonstrate that a firm’s strategy still may not succeed. The competitive context is always changing and, as the Cynefin framework tries to demonstrate, a stable, linear context can swiftly evolve into an unstable, non-linear context. The medium sized fast-moving consumer goods firm, Case 2, was once a pioneer in establishing new product characteristics, but is now faced by the rising levels of competition from incumbents and new entrants copying its approach. The heavy manufacturer, Case 6, faces global macro-economic pressures that influence its competitiveness. This competition from emerging economy markets can provide standardised product to world markets but not provide the same environmental or social sustainability performance criteria required by the regulatory context or consumer values in the West. This means that unless there is a standardising of ecological or social criteria across global markets, such measures can add to bottom line costs, undermining competitiveness.

This returns to the early debate on the alignment developed within the strategic management field, which concerns the link between strategy and the external environment (Henderson and Venkatraman, 1993; Venkatraman and Camillus, 1984). The conclusion from this is that internal alignment or fit for the PMM does not necessarily ensure success. It may be a necessary condition of organisational longevity, but it is not a sufficient condition. If the PMM is not well-aligned, the performance will be suboptimal at best. Hence, PMM must still align with strategy, while strategy aligns with the external environment. The lines of alignment shown in Figure 3 are validated by the research.

**SOSCM and PMM**

SOSCM as PMM is a new undertaking for many organisations and a relevant area for future research in its own right. This study has focussed on SOSCM as an implementation of PMM primarily from the decision theory perspective via the Cynefin framework. One general proposition to emerge is that by virtue of its novelty to many organisations, SOSCM represents a challenge for both PMM and strategy.

Focussing on the PMM aspects, a failure of strategic alignment, or lack of fit, can be found in terms of internal alignment. This concerns the extent to which there is a disconnect between the strategic direction of a firm sought by senior management, and the ability of operational-level PMM processes to align. Strategy is a case of external alignment. Here there is a challenge faced by many organisations as to how ecological and ethical issues are aligned or misaligned with their core strategy. For some businesses, external environmental changes in customer expectations or legislative change can challenge the core strategy. Firms dependent on fossil fuel extraction for instance, or changing attitudes to commodities such as palm oil, may need a change to the strategy of the business, or resistance to the new demands being exerted, commonly via lobbying against regulation or the rhetorical communications work of greenwash.

An additional finding from the data was in relation to DL, SOSCM and ownership type. In Cases 2 and 3, different ownership types placed less pressure to have clear and unambiguous economic Domain 2 justification for any SOSCM activity. Case 4 by contrast used a clear quantitative economic rationale to justify carbon reduction measures, but the volatility of oil price undermined this economic argument shortly after the interviews were conducted. This may mean that economic decision making for issues such as reducing carbon emissions are misplaced, and instead moral decision making is more appropriate.

Both firms in this study practicing this form of decision making as part of their DL (Cases 2 and 3) were privately owned, not publicly listed on stock markets. This prompts a future research question on the influence of ownership type on DL. The hypothesis would be that firms under private ownership may find it easier to establish values-based DL than
those listed on major stock markets. Given the latter’s need for quantitative, Domain 2 accountability to unknown investors challenges to this form of DL may represent a barrier to addressing the issues of sustainable and responsible business.

The nature of knowledge – the extent to which something is known or certain – is central to both Cynefin and the performance alignment matrix (as shown in Table I). Various aspects of decision theory (indeed, various decision theories), including the nature of advanced quantitative analytics – “big data” or “artificial intelligence” – have been subject to extensive consideration in academic literature dating back several decades. So too has the area of ethical decision making, and the role of judgement, intuition and moral principles. One core message of decision theory is that any form of decision support system, model, set of principles or rule of thumb, is merely a tool that the decision maker uses to help guide or inform them; the responsibility for the decision taken still remains with the decision maker’s.

Seen as a framework within decision theory, Cynefin places greater attention on the role of perception, and the significance of organisational culture as a shaper of that perception for decision-making via its DL. The performance alignment matrix, by contrast, addresses “outcomes and solutions” as problems to be solved rather than as decisions to be made. By examining the real-world case studies, a body of empirical data has been collected that seeks to understand how organisations relate to their external environment and how their PMM and strategy align. As Cynefin is a sense-making framework that addresses both predictable external contexts and unpredictable ones characterised by complexity and volatility (where cause and effect cannot be determined in advance), it also offers insight into the issue of complexity in PMM.

Looking at the use of PMM systems that concern the application of SOSCM into organisations, the experience of changes in PMM, changes in strategy and changes in external environment have been subject to context-rich, in-depth exploration. Theoretical explanation of this can then be stated in terms of the Cynefin framework. Underlying assumptions of linearity or predictability are the key. For instance, the performance of energy consumption as part of programmes to improve energy efficiency is an instance of PMM. This was commonly framed as a driver of cost saving on the basis of a common assertion (found in interviews in Cases 4-6) that energy prices would only rise.

A similar example of SOSCM is also seen in Case 6 where LCA fails as a process to drive improved ecological performance. Merely having a certificate scores a point as a Domain 1 activity in supplier selection decision processes. Despite trying to generate accurate data, in practice no Domain 2 structured-complicated decision analysis could be developed to establish a clear, rational calculation of the benefit of one product over another. Instead, the process is actually a Domain 3 context, where there is a high level of ambiguity and context specificity. To make LCA amenable to structured decision making, such as supplier selection decisions based on environmental performance, would require a greater structuring via standardization of methodologies, as well as other means of determining valid comparability, that may not be possible (Q6.2).

By contrast with the performance alignment matrix, the Cynefin framework makes the different contrasts more explicit and provides deeper explanation as to why in terms of knowledge management and sense-making. It thus provides an extension or alternative perspective on the same phenomenon and categories described in the performance alignment matrix.

The main contribution provided by the research presented in this paper is illustrating the role of awareness of an organization’s context. Conditions of volatility, uncertainty, complexity and ambiguity can mean that strategy is forced to change quickly, and that PMM must then change equally fast at strategic, tactical and operational levels. The example of a client making write-downs because of the impact of the balanced scorecard
as a PMM in Case 5 (again, ultimately caused by volatility in real estate, post 2008),
demonstrates the huge and negative strategic impact that can result from a brittle PMM
that fractures in relation to strategy and environment. Organizational awareness,
attention, or sense-making are central to this.

While this paper has used the Cynefin framework to elaborate existing theory in PMM,
notably the performance alignment matrix in the work of Melnyk et al. (2014), the wider
academic field of decision theory, related to the themes in both Cynefin and the performance
alignment matrix, provides deep foundations from which to further develop theory.
Returning to development of the alignment concept in strategic management (Chorn, 1991;
Henderson and Venkatraman, 1993; Venkatraman, 1989; Venkatraman and Camillus, 1984)
provides further opportunities to elaborate the theory of fit in PMM, potentially revisiting
the roots of the debate in contingency theory (Hickson et al., 1971).

Conceptual synthesis (Denyer et al., 2008; Tranfield et al., 2003) may extend beyond the
exploration of Cynefin and the performance alignment matrix here. The concept of
dynamic capabilities, for instance, maintains that changing core capabilities in a swift and
responsive manner is vital for strategic success (Eisenhardt and Martin, 2000; Teece et al.,
1997). In the operations and supply chain management context, the hybrid concept of
“leagile” refers to the balance between lean and agile where the former suits stable, regular
process flows in mass production, against the operational flexibility required for dynamic
capabilities to stay responsive to changing customer needs and other external forces
(Purvis et al., 2014). These parallels and precursors further validate the concept of balance
between different contexts (suddenly changeable and unpredictable, requiring
responsiveness, or stable and predictable, and hence mathematically optimisable such
as via statistical process analysis).

This distinction echoes longstanding discussions in business and management
scholarship. These include the development of decision theory by Herbert Simon (1947,
1957, 1959, 1960, 1962, 1972, 1977, 1978); also, the critique of classical management science
contexts are referred to as hard systems, and unstructured are soft systems (Figure 2),
characterised by an inability to determine the variables; similarly, the concepts of tame
problems, messy problems and wicked problems (Rith and Dubberly, 2007; Rittel and
Webber, 1973) address the distinction later defined in terms of predictable, ordered
simplicity and unpredictable complexity in Cynefin. Similar precursors also exist in the
philosophy of science, such as Weaver (1948) “Science and Complexity” and Popper (1965)
“Of Clouds and Clocks”. These precursors are examples of reciprocal synthesis, where the
same phenomena are conceptualised using different terminology and from different
scholarly perspectives. This range of literature helps to validate the nature of the Snowden’s
Cynefin framework (and indeed Melnyk’s performance alignment matrix and Chorn’s
typologies for strategic alignment).

**Implications for research and practice**

As the first paper to consider the Cynefin framework in relation to operations
management and PMM, it has provided insight into the link between organisational
culture (the DL) and sense-making. This answers the call of Melnyk et al. (2014) on the
need for greater understanding of social factors, organisational culture, complexity,
autopoiesis and uncertainty in PMM theory, and the call for this special issue. In
addition, the relatively nascent area of SOSCM and PMM has been explored, the
newness of which is a clear driver of potential misalignment with existing PMM
systems and strategic objectives.

Making appropriate adjustments to PMM over time based on turbulent and variable
external environments and shifting between four domains of the Cynefin framework
could prove an effective way of reformulating the organisation strategy, promoting intended behaviour and organisational learning. An organisation that is aware of the different DLs presented in Cynefin may be better prepared for change. Yet, if an organisation is in a stable environment, such investment may appear as a suboptimal use of resources. A broader point is that the mind-set of managers and their organisational culture (their institutional schema for sense-making) is often very strong; the definition of what Cynefin is, “a Welsh word that signifies the multiple factors in our environment and our experience that influence us in ways we can never understand” (Snowden and Boone, 2007).

Whilst strong and committed leadership may be essential for reshaping organisational culture in preparation of diverse and turbulent environment (as in Domains 3 and 4), expectations about the locus of control are central to this theory. To advance theory around Domain 3 PMM as a means to address emergence and complexity, an open source approach to PMM might need to be considered. This is suggested in part by continuous improvement (Slack et al., 2009) where a bottom-up approach to change is taken. It is also formalised in the theory of positive deviance or anonymous employee feedback systems (Spreitzer and Sonenshein, 2004). These may suggest new, more responsive approaches to PMM, recognising bounded rationality as an issue for centralised authority, and capturing the wisdom of crowds as a form of decentralised emergence.

Combining decision theory with the topic of PMM draws attention to where decision making should lie. The acquiring of performance data is intended to enable managers in positions of authority to make decisions. Under the traditional model of organisational authority, the command-and-control approach, it is the senior management who acquire the data and then push the decisions downwards through a rigid hierarchy, where middle management are empowered to make decisions on the basis of analytic justifications, and “shop floor” decision making is standardised. However, this model does not reflect the need for emergence to address responsiveness under conditions of complexity. The Cynefin framework provides a means to map these different types of context and their responses for decision makers. However, the next stage for developing PMM would be how to develop emergent performance measurement and management processes.

This relates to how PMM connects to internal organisational structure in order to produce better run organisations. Organisations applying an integrated Cynefin-aware PMM system may have a dynamic policy/strategy deployment process in place where changes in the external environment can be swiftly reflected internally in the organisation. Reflecting these changes in the organisational strategy at the business level would involve interaction with the PMM system at tactical and operational levels. Such an organisation would promote empowerment, flexibility, skill development at every level and at the same time have people able to probe, sense and respond to external uncertainty. Examples of organisations that are exemplars in dynamic capabilities, or who combine both lean and agile characteristics may provide significant cases for future research. Where lean tools are capable of driving improved performance in a stable environment, organisations that can dynamically combine lean with agile when faced with unpredictable change, or persistent uncertainty in the external environment, provide valuable examples of how managers cope with what is knowable or unknowable. How managers best respond as a result, including how the organisation as a whole responds, through organisational culture, strategy formulation and PMM processes, are the central questions.

The data gathered here have encountered instances of organisations with different types of culture and PMM processes. The concept of DL is a useful way to extend theory around organisational culture into the PMM literature, particularly with regard to how the topic of changing a DL is recognised as a major challenge (Case 4). Furthermore, using Cynefin as a
way to deeper the theoretical concept of DL is a helpful way to understand the relationship between organisational culture and the mindset of people within a given industry, and the potential ways in which this helps deepen understanding of the organisational culture in relation to PMM. Clearly, the limitations to this research exist in terms of the scope of the cases studied and the limited number of industrial sectors included. Further conceptual development is possible, and the bridges to previous work in business and management scholarship suggest that deeper conceptual reviews of the literature in parallel fields such as strategic management may be beneficial.

Future research is needed to highlight how the DL of different industries can be translated into more responsive PMM. For example, the restaurateur culture in Case 3 was deliberately created in a multi-national enterprise, in contrast to a more mechanised, analytic-driven culture that may typically be seen in the fast food sector. Similarly, the principles-based culture developed for the construction site is a marked contrast to that of organisations operating in the controlled and stable conditions of a manufacturing plant. To what extent are these cases typical of their sector, or was their alignment fortuitous?

What the research presented here has shown, are the instances where these DLs have encountered instances of misalignment with PMM systems. In line with the processes presented in the works of Eisenhardt (1989) and Ketokivi and Choi (2014), the purpose of such a method is to better understand the concepts under investigation by deep engagement with organisational practice. Decision theory and the Cynefin framework is thus found to elaborate existing conceptualisations of unpredictability and complexity and subsequent problems of misalignment in PMM. New research is thus needed to further investigate the potential for bringing PMM and decision theory together in order to better understand how to improve the performance of organisations operating in an increasingly fast-changing and unpredictable world.

References


Melnyk, S.A., Hanson, J.D. and Calantone, R.J. (2010), “Hitting the target ... but missing the point: resolving the paradox of strategic transition”, *Long Range Planning*, Vol. 43 No. 4, pp. 555-574.


**Further reading**


Appendix

Ref. no.  Representative quote

Q1.1  “We are an engineering company. We are more focussed on quality and assurance, and that’s how we deal with our suppliers […] There’s not a lot of revolution […] The rate of revolution is very low” (Financial director)

Q1.2  “We wouldn’t have any say on the components that were used to give us the product. We have no input there […] I would say we have zero influence to be absolutely clean. We have zero influence as to the material they are going to be using and the process they are going to be using” (Financial director)

Q1.3  “In the process we use a lot of chemicals […] harmful […] to the planet. The difference is how well you control that process […]. Quality, performance and service should be driving that anyway. If they are very good in quality, service and availability, they are most likely to be efficient in maintaining the equipment, therefore they should have less wastage” (Financial director)

Q1.4  “One issue that I’ve got is when I look at my organisation, it’s not in the DNA to do CSR. Of the group or the individual” (Financial director)

Q1.5  “people understand the words sustainability and corporate responsibility now […] They get the fact that we need to be cleaner technology. They get some stuff around carbon goals […] but don’t forget we are an engineering firm so people are very linear in their thinking. Whether they are a design engineer or one of the quality guys […] people are incredibly linear and don’t necessarily get off their islands […] they’re not used to doing that” (Corporate social responsibility manager)

Q2.1  “As a values-based business, we recruit on the basis of ‘will people be able to fit in and […] adopt and support our values […]’. We have a [regular] review session that checks-in on how are the values living in the office […] we deliberately don’t have a leader on each of the values as the idea is that everyone in the team is fostering, and working to drive, those values.” (Managing director)

Q2.2  The way the values come out day-to-day is that […] you’ll hear people use them […] When we’ve got a problem or an opportunity, rather than say […] ‘get the sales team together’ and say, “there’s a retail opportunity”, actually, "who’s around who can help input”. Whether that’s the team PA, the supply chain manager, the marketers, the finance person, or the [retailer] account manager” (Managing director)

Q2.3  “it’s the whole David vs. Goliath thing, that we don’t have the same resource as [rival firm #1] and [rival firm #2] and [rival firm #3] have, so we’re going to have to be fast and nimble and use the resources we do have to maximum effect” (Managing director)

Q3.1  “The culture takes a bit of getting used to […] It’s a big company where it’s got a very small company feel […] there’s the mindset for a start. You don’t have to go in and battle different people for different reasons. You don’t have to go and battle with the FD because he’s got a different outlook, about money, or the Ops Director […] it’s just accepted that I’m doing it because it’s the right thing to do […] they will try things and it’s ok to fail. But you don’t know until you try” (Energy manager)

Q3.2  “No one is saying to me, ‘where’s your compliance monitoring’ […] we need to go through loads more boring processes, that move us extremely slowly and just get us a figure at the end of a year’s work’ - and then we don’t know what to do with that figure […] if I get an idea that’s got legs, they’ll go for it. But if I start giving them formal structures, and say ‘this is where we are at on these formal structures, we almost slow our own progress […] I remember one of the regional managing directors, who at the time I thought was just a hard-nosed operator […] said, ‘I just want to be able to sleep easy in my bed at night’” (CSR manager)

Q3.3  “We get clobbered with everything. CRC [Govt ‘Carbon Reduction Commitment’ scheme], ESOS [Govt ‘Energy Savings Opportunity Scheme’], which is coming in this year […] We have to report on our annual consumption of gas and electricity, then we have to pay an amount per ton of CO₂ emitted […] we have to audit […] and produce reports and that’s it really” (Energy manager)
Q4.1 "[...] we have all sorts of delegated levels of authority within [the company]. We have more risk process checkers, checkers checking checkers than you would believe [...] really risk averse internally [...] it all went very control focused [...] and we are reaching a situation where this very risk-averse position is unsustainable. So we need to rebalance and it's that rebalancing, I think, that's led to a more principles and values-based approach [...] It's not a compliance-based, you must follow this set of rules. It's about equipping people with the principles that they need and the values to make that decision independently, without having to rely on a system of rules [...] the more rules you put in place, the less independent thinking there is within an organisation." (Supply Chain Director)

Q4.2 "Our fuel costs are second behind payroll [...] we run a huge fleet of vehicles [...] we've got to think outside the box because fuel costs will continue to rise [...] and it's not just the fuel costs of the fleet, you know, we spend a lot of money on utility costs - gas, electric heating oil; it's a huge spend for the business [...] let's stretch our thinking on this because energy costs ain't gonna be dropping". (Logistics supplier, Head of environment)

Q5.1 “[Construction] projects are a series of problems that are to be solved. If it was easy to build them, you wouldn't need a principal contractor, because you just tell the subbies [sub-contractors] to turn up and send them a little list and when they should turn up. They are complex and they are one-off and that's difficult [...] We've got a really good PM [project manager] who's done jobs like this before, and will do jobs like this again, but on every one of his jobs he's got a massive learning curve [...] That can't be under-estimated. The learning curve is dramatic in each project [...]” (Regional director)

Q5.2 "Health and safety drives it. Construction kills a lot of people every year - an awful lot of people. And that's just not good enough. We are going to build things. We don't want to kill people doing it. So that's the driver” (Contractor, Regional director)

Q5.3 "The guys that manage these sort of projects, they're just fantastic to watch - to see them do their job. They can juggle hundreds of things at the same time, and they've almost built this a hundred times in their head. They've got to get that deep into it, so when a sub-contractor comes in and asks them the silly question, they can answer them straight away and push them back out the door” (Sustainability manager)

Q5.4 "BREEAM is only a tool and there's various tools out there for various different things. It's just a tick box. Things like BREEAM actually are quite misguided as well. Building regs tend to be more solid and drive performance to be what we want to get from places. BREEAM is a little bit more airy fairy to be honest". (Logistics supplier, Head of environment)

Q5.5 "We have a set of KPIs we use, but to be honest, they are in and out a little bit. The balanced scorecard is a pretty impressive thing, but it is a pretty impressive thing when you are doing the same thing [...] when it is aligned. It's more difficult when you are dealing with a series of unique projects [...]" (Regional director)

Q6.1 "It's high spec, specialist [engineering] where I have a team of engineers working for me who do the calculations. They rock up and, say I want an architectural facade on this, and he and his team would tell you exactly what you needed – calculate it all out. There are various models [...] and all that stuff” (Commercial director)

Q6.2 "We don't use [life cycle analysis] as a tool for improvement at the moment. We use it in terms of [external auditing]. We are struggling to understand the definitions, particularly when our customers will define life cycle analysis to suit their own need and then use that as a comparison [...] Our competitors [...] they've got a life cycle that is based on slightly different starting points, slightly different definitions to make their system advantageous over ours. So, it's difficult to compare like with like. I think our customers would love to be able to use life cycle analysis as a selection tool as part of their discussions. Until we actually define what the full standards of that are and the initial unit of measured definitions that go into that, it becomes a little bit difficult to do like for like comparisons. I guess there's a need for standardization” (Operations and supply chain director)

Q6.3 "[...] within a tender process [...] the price dominates. So you can be high in technical [...] highly sustainable and still lose out or not get a fair value for that because there is a lot of emphasis on price” (Commercial director)

(continued)
Corresponding author
Anthony Alexander can be contacted at: fm10aa@mail.wbs.ac.uk

<table>
<thead>
<tr>
<th>Ref. no.</th>
<th>Representative quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6.4</td>
<td>“It is so very difficult to make a long term decision [...] The changeable nature of the market is very difficult to deal with at the moment. It’s almost paralyzing [...] You ask any energy buyers across the country and they’ll say the biggest problem is legislation changes. We don’t know where we are. How can we make a decision on legislation that could change tomorrow on a moment’s notice” (Energy buyer)</td>
</tr>
<tr>
<td>Q7.1</td>
<td>“[what we do is] very simple and very standard”. (Communications director)</td>
</tr>
<tr>
<td>Q7.2</td>
<td>“When we went through the KPIs and we analysed each and every one trying to clearly understand exactly what we were trying to measure [...] what we came up with was a detailed list of KPIs that whether you had a sustainability program or not, it would be the right things to monitor, measure and focus on because it’s the right way to run a business”. (Chairman)</td>
</tr>
<tr>
<td>Q7.3</td>
<td>“[...] the governance around it basically is quite structured. We have our sustainability agenda that’s governed by our board committee [...]. Then we have sponsored group champions, and site champions which work with the group champions, to deliver on the KPIs; and underlying all that is the non-financial KPIs” (Communications director)</td>
</tr>
</tbody>
</table>

Table AI.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Volume 38 Number 11 2018

International Journal of Operations & Production Management

Performance measurement and management: theory and practice
Guest Editors: Mike Bourne, Steven Melnyk and Umit S. Bititci

2009 Editorial board
2010 Guest editorial
2022 The relationship regulator: a buyer-supplier collaborative performance measurement system
Vieri Maestrini, Veronica Martinez, Andy Neely, Davide Luzzini, Federico Caniato and Paolo Maccarrone

2040 The impact of supplier performance measurement systems on supplier performance: a dyadic lifecycle perspective
Vieri Maestrini, Davide Luzzini, Federico Caniato, Paolo Maccarrone and Stefano Ronchi

2062 Missing performance management and measurement aspects in performance-based contracting: a systematic process-based literature analysis of an astonishing research gap
Andreas H. Gläs, Florian U. Henne and Michael Essig

2096 Operations flow effectiveness: a systems approach to measuring flow performance
Mohamed Afy-Shararah and Nicholas Rich

2124 The performance effects of complementary management control mechanisms
Henrik Nielsen, Thomas Borup Kristensen and Lawrence P. Grasso

2149 Controlling lean manufacturing in multidivisional organisations: highlighting local interests and constraints
Martijn Pieter van der Steen and Sandra Tijmena

2169 The use of management control and performance measurement systems in SMEs: a levers of control perspective
Biljana Pešalj, Andrey Pavlov and Pietro Micheli

2192 Predicting performance – a dynamic capability view
Marcus F. Hasegan, Sai Sudhakar Nudurupati and Stephen J. Childe

2214 A decision theory perspective on complexity in performance measurement and management
Anthony Alexander, Maneesh Kumar and Helen Walker

ISBN 978-1-78973-199-6

www.emeraldinsight.com/loi/ijopm