



Heriot-Watt University
Research Gateway

Best practices, performance advantage and trade-offs

Citation for published version:

Sousa, R, Camanho, AS, Silva, MC, da Silveira, GJC & Arabi, B 2023, 'Best practices, performance advantage and trade-offs: new insights from frontier analysis', *Journal of Productivity Analysis*.
<https://doi.org/10.1007/s11123-023-00714-y>

Digital Object Identifier (DOI):

[10.1007/s11123-023-00714-y](https://doi.org/10.1007/s11123-023-00714-y)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Productivity Analysis

Publisher Rights Statement:

© The Author(s) 2023.

General rights


Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Best practices, performance advantage and trade-offs: new insights from frontier analysis

Rui Sousa¹ · Ana S. Camanho² ² · Maria Conceição Silva¹ · Giovanni J. C. da Silveira³ · Behrouz Arabi⁴

Accepted: 10 November 2023
© The Author(s) 2023

Abstract

There are still important theoretical and empirical gaps in understanding the role of best practices (BPs), such as quality management, lean and new product development, in generating firm's performance advantage and overcoming trade-offs across distinct performance dimensions. We examine these issues through the perspective of performance frontiers, integrating in novel ways the resource-based theory with the emergent practice-based view. Hypotheses on relationships between BPs, performance advantage, and trade-offs are developed and tested with stationary and longitudinal (recall) data from a global survey of manufacturing firms. We use data envelopment analysis, which overcomes limitations of mainstream methods based on central tendency. Our findings support the view that BPs may serve as a source of enduring competitive advantage, based on their ability to lead to a heterogeneous range of dominant and difficult-to-imitate competitive positions. The study provides new insights on contemporary debates about the role of BPs in generating performance advantage and how practitioners can sustain internal support and extract benefits from them.

Keywords Manufacturing strategy · Data envelopment analysis · Best practices · Performance frontiers

1 Introduction

The implementation of operations management best practices (BPs) such as quality management, lean and new product development (NPD) has been considered a central strategy to improve operational performance (da Silveira and Sousa 2010) and to help mitigate trade-offs between dimensions such as cost, quality, delivery and flexibility (Schmenner and Swink 1998). However, understanding the role of BPs in generating performance advantage is still incomplete and is by no means a closed issue (Hitt et al. 2016).

First, the long-standing debate on whether and how BPs can provide sustainable advantage—since they are not proprietary—is very much alive. This debate has recently resurfaced with the practice-based view of the firm (Bromiley and Rau 2014, 2016). The practice-based view is a complementary theoretical perspective to the resource-based theory regarding firm performance (Hitt et al. 2016). Contrary to the resource-based theory, the practice-based view proposes that implementing practices which are even in the public domain can provide significant and enduring performance benefits to operations.

Second, recent work has questioned BPs ability to support sustained benefits. Evidence suggests that established programs such as lean and quality management may be losing management interest and providing less than expected performance benefits (Fundin et al. 2018; Schonberger 2019). However, these difficulties seem to be more common with mature programs and at higher levels of operational performance (Netland and Ferdows 2016) and may be related to the inability of firms to continuously adapt existing programs to a changing context (Spear 2004). Such findings need to be reconciled with the practice-based view and the numerous empirical studies which have suggested positive relationships between BPs and operational performance.

✉ Ana S. Camanho
acamanho@fe.up.pt

¹ Universidade Católica Portuguesa, Católica Porto Business School and CEGE, Porto, Portugal

² Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

³ Haskayne School of Business, University of Calgary, Calgary, AB, Canada

⁴ INESC TEC, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

Finally, studies exploring relationships between BPs and performance advantage have suffered from limitations with performance concepts and empirical designs. Although studies have often shown a positive impact of BPs on individual performance dimensions (Mackelprang and Nair 2010), there is less understanding of BPs potential to provide distinctive, difficult-to-imitate competitive positions. A firm's operational advantage may depend less on scores in individual dimensions than on its overall (multidimensional) performance vis-a-vis their competitors (Schmenner and Swink 1998). In addition, the extent that BPs can drive simultaneous improvements across performance dimensions (mitigating trade-offs) may depend on a firm's position relative to the performance frontier (Rosenzweig and Easton 2010). Specifically, as a firm gets nearer the frontier, BP programs may yield diminishing returns making it more difficult to improve on one performance dimension without weakening another (Schmenner and Swink 1998; Rosenzweig and Easton 2010). Few studies seem to have tested this hypothesis by estimating distances to performance frontiers. We still need more suitable conceptualizations of multidimensional performance advantage, and empirical designs that capture distances to PFs.

This study sets out to address these gaps and provide new insights about the role of BPs in generating operational performance advantage and mitigate trade-offs in operations. To do so, we draw on the theoretical and empirical perspectives of frontier analysis (Charnes et al. 1978; Devinney et al. 2010). Theoretically, as we argue later, frontier analysis allows the integration of the resource-based theory and practice-based view perspectives of BPs in operations. Empirically, frontier analysis appears more able than methods based on central tendency to capture the multidimensional nature of performance advantage. Besides, it allows the estimation of a firm's distance to the frontier, which may be explained by the development of distinctive capabilities (Devinney et al. 2010; Rosenzweig and Easton 2010).

Based on frontier analysis, we develop hypotheses on the role of BPs in leveraging operational advantage and mitigating trade-offs. Hypotheses are tested using data envelopment analysis (DEA) (a mainstream frontier method) based on a large international survey of manufacturing firms, including stationary and longitudinal (recall) data.

2 Theoretical background

2.1 Frontier analysis

Frontier analysis allows the comparison of firms over multidimensional performance profiles, based on the notion of dominance (Devinney et al. 2010). Within a set of comparable firms (e.g., belonging to the same industry), a firm is said to “strong-form dominate” (outperform) a set of

other firms when it exhibits performance on all dimensions that is superior to all other firms. Because in practice few firms achieve this superiority, the related concept of “linear dominance” is more frequently used. A firm linearly dominates (from now on, “dominates”) another set of firms when there is no linear combination of those other firms that can do better across all dimensions of performance.

Applying frontier analysis to a group of firms assumes they are subject to comparable contextual factors that result in a similar performance potential; contextual factors are those that are outside the control of firms or can only be changed in the long term (Dyson et al. 2001). Frontier analysis works through the construction of the frontier, where “best-in-class” firms are located—firms which, by effective managerial action, have fulfilled their performance potential. The approach allows for the measurement of distances of individual firms to the frontier, which signals performance improvement potential for non-frontier firms.

2.2 Best practices, performance advantage and trade-offs in operations

In operations management, studies have converged to two major sources of operational performance: the development of manufacturing capabilities (anchored on the resource-based theory) and the adoption of BPs (anchored on the practice-based view) (da Silveira and Sousa 2010).

The resource-based theory states that the development of organizational capabilities—routines or bundles of routines oriented to specific objectives—is a key source of competitive advantage (Winter 2003). Recent advances in resource-based theory research have focused on the development of dynamic capabilities and their deployment in the market to achieve advantage (Sirmon et al. 2011; Hitt et al. 2016). Dynamic capabilities are the “routine-changing routines” (or meta routines) that enable a firm to change—either radically or incrementally (Helfat and Winter 2011)—operational routines related to products and processes (Winter 2003). They “confer upon an organization's management a set of decision options for producing significant outputs of a particular type” (Winter 2003, p. 991). In Operations Management, the resource-based theory has been supported by studies which showed that the development of manufacturing capabilities that are difficult to imitate, such as learning, can lead to improvements in individual dimensions of performance (Dabhilkar and Bengtsson 2008; Peng et al. 2008; Anand et al. 2009; da Silveira and Sousa 2010).

The practice-based view defines BPs as sets of activities (or routines) which are widely known and disseminated globally, the execution of which is expected to lead to better performance (Bromiley and Rau 2014). In operations management, BP programs usually comprise bundles of individual practices within a thematic label (Shah and Ward 2003),

such as quality management, lean or NPD. Seemingly contrary to the resource-based theory, the practice-based view advocates that implementing practices available from the public domain can at least in part explain variations in performance across firms (Bromiley and Rau 2014). This view is supported by many studies which found positive correlations between BPs and improvements in individual dimensions of performance (see reviews by Sousa and Voss (2002) and Mackelprang and Nair (2010)).

We argue that a key link between these two views is the notion that the implementation of BPs can lead to operations-based dynamic capabilities (Peng et al. 2008; Anand et al. 2009; Linderman et al. 2010; Bititci et al. 2011); operations-based dynamic capabilities have been defined as bundles of interrelated routines oriented to specific objectives, including improvement (incremental change) and innovation (radical change) at the manufacturing plant level (Peng et al. 2008). This is because (i) BPs correspond to the implementation of meta routines in the manufacturing organization (for example, BPs associated with continuous improvement lead to the change of day-to-day production routines); (ii) the routines lead to change (improvement) in products (a core aspect of NPD, for example) and processes (a core aspect of quality management and lean, for example) (Peng et al. 2008); (iii) the implementation of BPs involves organizational learning (Anand et al. 2009; Linderman et al. 2010).

Thus, anchored on this notion, the practice-based view and the resource-based views can offer complementary explanations for the relationship between the adoption of BPs and performance advantage (Peng et al. 2008; Anand et al. 2009; Bromiley and Rau 2016; Hitt et al. 2016) (section “BPs and operational advantage” provides additional elaboration on this relationship). We submit that a frontier analysis framework can reconcile the core principles of resource-based theory and practice-based view and increase our understanding of the links between BP adoption and operational advantage, as follows.

In frontier analysis, empirically, the distance to the performance frontier can be interpreted as the gap between the distinctive capabilities of a firm and of their peers (Devinney et al. 2010). Thus, by focusing on the comparative advantage of frontier firms, frontier analysis fits to the resource-based theory view of dynamic capabilities as sources of sustainable and hard-to-replicate advantages (Helfat and Winter 2011). Frontier analysis also aligns with the notion of the practice-based view that, by exploiting the implementation of BPs in different ways (see section “BPs and operational advantage” for further details), firms have managerial leeway to select a location in the multidimensional performance space that matches their competitive environment and provides advantage; however, their choice space is bounded by the frontier (Devinney et al. 2010). Moreover, frontier analysis captures the practice-based view

notion that firms benchmark each other’s practices and performance, and then engage in adoption of rivals’ practices to reduce discrepancies if they are below the benchmark (Hitt et al. 2016). Managers will compare their performance to the factored performances of various actual and potential firms, rather than a single competitor (Devinney et al. 2010). Frontier analysis aligns with this logic by comparing firms with a frontier made up of not only actual peers, but also notional firms that might enter the market. However, unless firms can continuously change (i.e., by developing dynamic capabilities), the advantage gained from the mere adoption of rivals’ practices quickly erodes.

We also submit that frontier analysis can overcome empirical limitations of studies on BP-based competitive advantage, as follows. Empirically, frontier analysis can provide fundamental advances over other approaches to studying the impact of BPs on performance advantage (e.g., Mapes et al. 1997; Lapré and Scudder 2004). Few studies established a clear link between BPs and unique capabilities leading to multi-dimensional dominating positions near or at the performance frontier (Lieberman and Dhawan 2005). Studies based on frontier analysis could lead to more valid evidence compared to correlation studies looking at performance dimensions independently, rather than the multidimensional competitive position of a firm relative to its peers (Devinney et al. 2010). In particular, we are not aware of studies that used the traditional set of operations performance dimensions (i.e., quality, delivery, flexibility, and cost) to estimate n-dimensional competitive positions and performance frontiers in a manufacturing sample.

Frontier analysis can also address key limitations of prior studies examining trade-offs in operations. There is a long-standing debate about which laws of cumulative capabilities and trade-offs prevail, and under what conditions (see for example the meta-analysis by Rosenzweig and Easton 2010). It has been suggested that the law of trade-offs becomes stronger, and the law of cumulative capabilities becomes weaker the closer the firm gets to the performance frontier (Sarmiento and Shukla 2011). However, many empirical trade-off studies have been limited by not considering distances between observations and their performance frontier, as well as by not using longitudinal performance data. By aggregating all units—far and near the frontier—in correlational analyses of performance dimensions using stationary data from cross-sectional samples, the studies could not test the existence of “frontier trade-offs” (Schmenner and Swink 1998; Sarmiento and Shukla 2011). Consistent with this, extensive reviews of trade-off research (Rosenzweig and Easton 2010; Chatha and Butt 2015) indicated that many more studies supported the cumulative capabilities model than the trade-off model.

That is not to say there have been no contributions in that direction. For example, even without measuring distance to a frontier, Netland and Ferdows (2016) found that the impact of lean programs diminished at higher levels of

performance, which is consistent with the proposition. In a similar vein, Schroeder et al. (2011) indicated that support to cumulative capabilities was contingent to distance to the performance frontier.

A frontier framework integrating multidimensional performance, distance to performance frontiers and longitudinal data might provide a more rigorous examination of trade-offs in operations.

3 Research hypotheses

We develop hypotheses of associations between BPs, performance advantage and trade-offs, informed by frontier analysis, resource-based theory, and the practice-based view. Following frontier analysis, they relate to a group of firms operating in comparable contexts. To overcome limitations of prior research, the hypotheses are testable via a mix of stationary and longitudinal (recall) data.

3.1 BPs and operational advantage

The implementation of BPs can lead to sustainable operational advantage in two ways. The first is anchored on practice-based view's four potential explanations for performance variation (use of specific practices, how the practices are used, interactions between practices, behavior of competitors) (Bromiley and Rau 2014). These are related to the content of BPs and can lead to a heterogeneous range of dominant (improved and differentiated) competitive positions (performance enhancement). BPs content (e.g. procedures, standards, and tools) can support direct improvements in multiple performance dimensions (Swink et al. 2005). For example, as indicated earlier, organizational integration between design and production may lead to simultaneous improvements in product flexibility and cost. In addition, idiosyncratic integrations between practices (Shah and Ward 2003) provide opportunities for competitive differentiation. This resonates with the resource orchestration approach, which has been argued to support core tenets of the practice-based view (Hitt et al. 2016). According to this approach, the way capabilities are bundled and coordinated can be a source of differentiation (Sirmon et al. 2011). In addition, the performance objectives of BPs can be redirected over time (Schroeder et al. 2002). For example, a continuous improvement meta-routine can be similarly employed to improve cost or flexibility by different firms or by a same firm at different points in time.

The second way BPs may lead to sustainable advantage is anchored on resource-based theory's focus on dynamic capabilities. These are related to the processes of BP implementation and can lead to competitive positions that are difficult to imitate. This in turn can result from two mechanisms. The changes and improvements in first-order routines

brought about by the implemented BPs involve second-order learning, associated with proactive change (Anand et al. 2009; Linderman et al. 2010). For example, interpreting charts from statistical process control may help not only to curb uncontrolled variance but also to eliminate its root causes, enhancing process knowledge. Such improvements are typically iterative and subject to causal ambiguity, leading to unique products and processes (Schroeder et al. 2002). Also, the content and objectives of BP programs often need to be continuously adapted, for example after changes in internal and external environments, or in competitive priorities. These adaptations can lead to unique, path-dependent improvement trajectories (Peng et al. 2008) which are hard to replicate.

These arguments suggest that BP usage should support improvements in multiple performance dimensions, enabling a firm to move closer to diverse segments of the performance frontier. We put forward the following hypotheses, testable with stationary (H1) and longitudinal (H2) data:

H1. The level of implementation of BPs is negatively associated with distance from the performance frontier.

H2. Over time, increments in the level of implementation of BPs are positively associated with the degree of simultaneous improvements in performance dimensions.

BP performance impacts are subject to diminishing returns related to the process of BP implementation, namely, due to learning curves and program adaptations. Learning curve theory posits that marginal improvements decrease with less improvement space remaining, as "low hanging fruits" become scarcer (Netland and Ferdows 2016; Zangwill and Kantor 1998). Besides, for multi-dimensional performance frontiers, enhancement costs may become so high that available resources can be exhausted on one critical dimension, leaving no reserves for other dimensions (Schmenner and Swink 1998). Concerning adaptation, as the improvement space shrinks by approaching the performance frontier, there may be fewer opportunities to extract unique performance gains from redirecting existing practices to new objectives. In other words, the choice of what "candidate" practices may still lead to effective improvements becomes narrower, requiring greater effort to make them successful. Thus, considering longitudinal data:

H3. Over time, the impact of increments in the level of implementation of BPs on the degree of simultaneous improvement of multiple performance dimensions is positively moderated by distance to the performance frontier.

3.2 BPs and trade-offs

Under the law of cumulative capabilities, a firm can simultaneously improve multiple performance dimensions, typically via BP usage (performance enhancement, da Silveira and Slack 2001). For example, a firm may achieve simultaneous improvements in flexibility and cost by developing closer integration between design and production. Under the law of trade-offs, improvements in one dimension will come at the expense of other dimensions, by reconfiguring operations choices (Schmenner and Swink 1998) (performance repositioning, da Silveira and Slack 2001). For example, a firm may decide to focus more on flexibility and less on cost by converting a flow-shop layout into a cellular layout.

We examine the occurrence of trade-offs relative to distance of a firm to the performance frontier. The closer a firm gets to its performance potential (or the performance frontier), the more resources it needs to attain further improvements, and the less it can improve at no expense to other dimensions (Schmenner and Swink 1998; Lapré and Scudder 2004; Sarmiento and Shukla 2011). As discussed in H1-H3, firms below the performance frontier may employ BPs for simultaneous enhancements on various performance dimensions (or enhancements on one dimension at no expenses of another). Thus, at one point in time, differences across the competitive positions of a group of firms below the performance frontier reflect both repositioning trade-offs and different levels of performance enhancements. For firms located on or near the performance frontier, enhancement is less feasible (H3), so that their competitive positions reflect overall performance potential and should reveal repositioning trade-offs (Schmenner and Swink 1998; Sarmiento and Shukla 2011). For example, the potential for cost performance of a firm employing a flow-shop layout should be higher than that of a firm employing a process layout. In other words, distinctive positions within this group would be more strongly determined by hard manufacturing strategy choices, which are subject to trade-offs. Thus, considering stationary data:

H4. Firms located on or near the performance frontier exhibit stronger trade-offs between performance dimensions than firms located far from the performance frontier.

4 Data

The analysis was based on data from the sixth edition of the International Manufacturing Strategy Survey (IMSS-VI). The survey was carried out in 2013–2014 in 22 countries,

targeting manufacturing companies from ISIC 25–30. Researchers initially contacted 7167 companies, of which 2586 agreed to participate. Valid responses were returned by 931 companies, representing 13% of the initial contacts. Responses were given at the business unit level, by the director of manufacturing or related function.

Participant countries (and valid responses per country) included Belgium (29), Brazil (31), Canada (30), China (128), Denmark (39), Finland (34), Germany (15), Hungary (57), India (91), Italy (48), Japan (82), Malaysia (14), The Netherlands (49), Norway (26), Portugal (34), Romania (40), Slovenia (17), Spain (29), Sweden (32), Switzerland (30), Taiwan (28) and USA (48). The distribution of firms by ISIC sectors was 282 (ISIC 25 - fabricated metal products, except machinery and equipment), 123 (ISIC 26 - computer, electronic and optical products), 153 (ISIC 27 - electrical equipment), 231 (ISIC 28 - machinery and equipment not elsewhere classified), 93 (ISIC 29 - motor vehicles, trailers, and semi-trailers) and 49 (ISIC 30 - other transport equipment). Respondents included a balanced mix of SMEs and large units with quantiles 155 (25%), 300 (50%) and 1000 (75%) employees. Non-respondent bias tests using publicly available market performance data obtained non-significant results.

The IMSS incorporates devices such as those proposed by (Podsakoff et al. 2003) to minimize common method bias (CMB). Written assurance was given to respondents that responses were confidential. The questionnaire was quite long with nine pages and hundreds of items, reducing significantly a respondent's ability to imply relationships between variables, or to maintain "consistency motif" (Podsakoff et al. 2003, p. 882) across responses. Questions used formal and pragmatic language, minimizing the suggestion of "socially desirable traits, attitudes, and/or behaviors" (Podsakoff et al. 2003, p. 882). We carried out statistical tests to check the effectiveness of these measures in our study. Following Harman's single factor logic (Podsakoff et al. 2003) we carried out two principal component analysis tests indicating what number of components described best our measures. Parallel analysis based on scree plot indicated that the best number of components was eight, while the minimum mean square error of prediction (Josse and Husson 2012) was achieved at six components. Finally, the use of DEA significantly mitigates the threat of CMB. This is because the efficiency scores computed for each firm—which are central in the analyses—depend not only on the firm's ratings on the four performance dimensions, but also on those of other firms in the sample.

4.1 Measures development

The analysis included two sets of variables. Operations performance included the plant quality, cost, delivery, and

flexibility (i) at present and (ii) as compared to 3 years before. Likewise, BPs implementation was measured (i) at present and (ii) as effort over 3 years. All measures were developed as reflective latent variables. We present in Appendix A the fieldwork questions, scales, and descriptive statistics. All responses were given on five-point scales.

Operations performance was operationalized by perceived results of the plant in quality (Q), cost (C), flexibility (F), and delivery (D). Following the research model, we used two different time specifications, namely (i) change over the past 3 years (QC, CC, FC, DC) and (ii) performance at the time of data collection (QP, CP, FP, DP). The four scales and their indicators had close correspondence to previous operationalizations in manufacturing strategy, e.g., Ward et al. (1998) and Avella et al. (2011).

BPs were estimated by (i) the perceived 3-year effort to implement (BPC), and (ii) the present level of implementation (BPP) of three classical manufacturing BPs: total quality management, lean management, and NPD. These practices are among the most widely researched and disseminated in the world (Schonberger 1996). Contemporary BP programs are quite broad and although quality management, lean and new product introduction practices may each have unique aspects, they also share many common traits (e.g., employee involvement). They are often implemented together as bundles since they have synergistic effects (Bevilacqua et al. 2017; Shah and Ward 2003).

Thus, following Koufteros et al. (2009), we considered second-order BPs factors (BPC, BPP) based on the respective first-order practice estimates. Cronbach's alphas of the second-order scales reflecting on eight first-order indicators were 0.86 (BPC, $n = 843$) and 0.87 (BPP, $n = 843$). The lean and TQM scales corresponded to BPs in previous manufacturing strategy studies such as Ahmed et al. (1996), while the NPD scale was consistent with the work by Paashuis and Boer (1997).

4.2 Measures validation

We used confirmatory factor analysis (CFA) to test the validity, reliability and unidimensionality of measures. The analysis was carried in R version 3.5.1 (R-Core-Team 2018) using RStudio version 1.1.456 (RStudio_Team 2015) and the lavaan package (Rosseel 2012). The CFA used maximum likelihood estimates, which is the preferred algorithm for cases with missing data (Peters and Enders 2002).

We tested two separate models, one for "present" estimates of performance and implementation (P-model), and another for performance "change" and implementation effort over 3 years (C-model). The initial P-model included the 20 first-order indicators of TQMP, LEANP, NPDP, CP, DP, FP and QP (see Appendix A). We also entered TQMP, LEANP and NPDP as second-order indicators of BPP. We

Table 1 P-model estimates

Variable	λ	SE	CR	AVE
TQM – present (TQMP)			0.846	0.647
Quality improvement and control	0.806	0.017		
Improving equipment availability	0.837	0.016		
Benchmarking/self-assessment	0.768	0.019		
Lean – present (LEANP)			0.742	0.595
Restructuring processes and layout	0.652	0.031		
Pull production	0.874	0.031		
NPD – present (NPDP)			0.789	0.554
Design integration	0.728	0.022		
Organizational integration	0.774	0.021		
Technological integration	0.731	0.022		
BPs – present (BPP)			0.844	0.646
TQMP	0.906	0.024		
LEANP	0.670	0.034		
NPDP	0.818	0.027		
Cost performance (CP)			0.726	0.571
Unit manufacturing cost	0.686	0.055		
Ordering costs	0.820	0.063		
Delivery performance (DP)			0.823	0.700
Delivery speed	0.806	0.022		
Delivery reliability	0.866	0.021		
Flexibility performance (FP)			0.735	0.582
Volume flexibility	0.800	0.029		
Mix flexibility	0.723	0.029		
Quality performance (QP)			0.797	0.663
Conformance quality	0.806	0.023		
Product quality and reliability	0.822	0.022		

used the cut-off criteria indicated by Hair et al. (2010) for $n > 250$ and $12 < m < 30$ ($CFI > 0.92$, $TLI > 0.92$, $RMSEA < 0.07$). Fit estimates were outside the recommended thresholds ($\chi^2/df = 4.759$, $CFI = 0.897$, $NFI = 0.873$, $TLI = 0.875$, $RMSEA = 0.072$ [0.066; 0.077]). Based on standardized loading estimates (λ) and the theoretical content of variables, we dropped four indicators including manufacturing lead time ($\lambda = 0.348$), procurement lead time ($\lambda = 0.322$), product customization ability ($\lambda = 0.644$) and new product introduction ability ($\lambda = 0.630$). The revised P-model had much improved fit ($\chi^2/df = 1.505$, $CFI = 0.990$, $NFI = 0.971$, $TLI = 0.987$, $RMSEA = 0.026$ [0.016; 0.035]).

As shown in Table 1, all average variance extracted (AVE) estimates were greater than 0.5, and all composite reliability estimates (CR) were greater than 0.7, indicating the scales had good convergence validity and reliability (Fornell and Larcker 1981; Bagozzi and Yi 1988).

As shown in Table 2 and based on the "most stringent" (Voorhees et al. 2016) AVE-SM method correlations

Table 2 Correlations between latent variables (P-model)

	TQMP	LEANP	NPDP	CP	DP	FP	QP
TQM – present (TQMP)	0.804						
Lean – present (LEANP)	0.607	0.771					
NPD – present (NPDP)	0.741	0.548	0.745				
Cost performance (CP)	0.243	0.179	0.219	0.756			
Delivery performance (DP)	0.397	0.293	0.359	0.210	0.837		
Flexibility performance (FP)	0.338	0.250	0.305	0.260	0.554	0.763	
Quality performance (QP)	0.452	0.334	0.408	0.149	0.621	0.548	0.814

AVE square roots are in bold across main diagonal

Table 3 C-model estimates

Variable	λ	SE	CR	AVE
TQM – effort (TQMC)			0.849	0.652
Quality improvement and control	0.793	0.017		
Improving equipment availability	0.843	0.015		
Benchmarking/self-assessment	0.785	0.017		
Lean – effort (LEANC)			0.738	0.590
Restructuring processes and layout	0.647	0.031		
Pull production	0.872	0.032		
NPD – effort (NPDC)			0.797	0.567
Design integration	0.743	0.021		
Organizational integration	0.796	0.019		
Technological integration	0.718	0.022		
BPs – effort (BPC)			0.829	0.621
TQMC	0.885	0.026		
LEANC	0.654	0.035		
NPDC	0.807	0.028		
Cost change (CC)			0.757	0.609
Unit manufacturing cost	0.818	0.042		
Ordering costs	0.742	0.040		
Delivery change (DC)			0.842	0.727
Delivery speed	0.831	0.019		
Delivery reliability	0.875	0.018		
Flexibility change (FC)			0.752	0.603
Volume flexibility	0.807	0.025		
Mix flexibility	0.745	0.025		
Quality change (QC)			0.836	0.718
Conformance quality	0.825	0.018		
Product quality and reliability	0.870	0.017		

between first-order scales were lower than the squared roots of their respective AVE estimates, which support discriminant validity (Fornell and Larcker 1981; Voorhees et al. 2016).

The initial C-model included the 20 first-order indicators of TQMC, LEANC, NPDC, CC, DC, FC, and QC (Appendix A), besides TQMC, LEANC and NPDC as second-order indicators of BPC. The model fit was below

acceptable ($\chi^2/df = 4.912$, CFI = 0.910, NFI = 0.890, TLI = 0.891, RMSEA = 0.071 [0.066; 0.076]) and some indicators had low λ estimates. To guarantee symmetry with the P-model, we dropped manufacturing lead time ($\lambda = 0.549$), procurement lead time ($\lambda = 0.490$), product customization ability ($\lambda = 0.655$) and new product introduction ability ($\lambda = 0.668$). The revised C-model had good fit ($\chi^2/df = 2.124$, CFI = 0.981, NFI = 0.965, TLI = 0.975, RMSEA = 0.038 [0.030; 0.045]), AVE and CR estimates (Table 3). All correlations between first-order factors were below their AVE square roots (Table 4). These results supported the scales validity and reliability (Fornell and Larcker 1981; Voorhees et al. 2016).

5 Hypotheses testing

In our study, we use DEA to build aggregate performance indicators (Cherchye et al. 2007; OECD 2008), comprising quality, delivery, flexibility, and cost. DEA is a mainstream frontier analysis method (Charnes et al. 1978; Devinney et al. 2010; Cooper et al. 2011), used to estimate the performance of units within their comparison group, represented by distance to the empirically derived frontier.

Using DEA is adequate for our purposes, for several reasons. As a non-parametric technique, it requires no a priori assumptions regarding the functional form of the frontier (Hjalmarsson et al. 1996). In addition, aggregate performance estimation with DEA does not require a priori (subjective) weight specifications for individual dimensions, which allows for flexibility in performance assessment and avoids subjective bias in estimating trade-offs.

5.1 Selection of a group of comparable firms

Consistent with the assumptions of frontier analysis, we studied a group of firms which were comparable in terms of relevant contextual variables over which the firm had little managerial discretion and that might affect performance potential (Sousa and Voss 2008; Dyson et al. 2001). Specifically, we accounted for country, industry and firm size,

Table 4 Correlations between latent variables (C-model)

	TQMC	LEANC	NPDC	CC	DC	FC	QC
TQM – effort (TQMC)	0.808						
Lean – effort (LEANC)	0.579	0.768					
NPD – effort (NPDC)	0.714	0.527	0.753				
Cost change (CC)	0.195	0.144	0.178	0.781			
Delivery change (DC)	0.318	0.234	0.289	0.353	0.853		
Flexibility change (FC)	0.339	0.250	0.309	0.371	0.601	0.777	
Quality change (QC)	0.408	0.301	0.372	0.323	0.673	0.614	0.848

AVE square roots are in bold across main diagonal

which are common performance covariates in the literature (Lieberman and Dhawan 2005; Sousa and da Silveira 2019).

The selection of the group within the IMSS sample was based on conceptual criteria, as well as on the need to maximize sample size. First, we focused on high-income countries as per the World Bank definition for 2012, the IMSS baseline year (gross national income per capita higher than or equal to 12,475 USD) (WorldBank 2019). These included the US, Canada, Japan, Taiwan, and countries in Europe (excluded countries were Brazil, China, Malaysia, India, and Romania). Firms in these countries have similarly high labor costs and are expected to have similar levels of access to production resources; in addition, they make up a large proportion of the IMSS sample (67.3%; $n = 627$). Within high-income countries, we focused on the largest group of firms with similar type of production technologies that could be formed, namely, the union of sectors ISIC 25 and 28 ($n = 387$, 61.8% of the high-income sample). These sectors are both associated with the manufacture of metal products (Zhou et al. 2009). Finally, within the high-income-ISIC25-28 group we focused on SMEs, using the 500-employee cut-off defined by the US Small Business Administration Office of Advocacy, employed in Census data (International Trade Commission 2010). SMEs are expected to have a similar access to production resources, and they made up most firms in this group ($n = 272$, 70.3%).

For this group, we addressed missing data on the four first-order variables involved in the estimation of performance frontiers (QP, DP, FP, and CP). Little's (1988) test on these variables was non-significant at the 5% level ($\chi^2 = 15.351$; $DF = 8$, $p = 0.053$). Therefore, data could be considered missing completely at random and listwise deletion of cases with missing data was acceptable (Fichman and Cummings 2003). This resulted in the deletion of 34 cases, so the usable sample consisted of 238 cases.

5.2 Estimation of performance frontier with DEA

Prior to testing the hypotheses, we estimated the performance frontier based on the four performance dimensions (QP, DP, FP, and CP). To that end, we employed the DEA-

based composite indicator model known as “Benefit-of-the-Doubt” (Cherchye et al. 2007) (model (1)). The frontier estimated in this way represents the best performance levels (performance potential) that a firm in a group could achieve by way of beneficial managerial choices.

$$\begin{aligned} \text{Max } E(y_{j_0}) &= \sum_{r=1}^s u_r y_{rj_0} \\ \text{s.t. } \sum_{r=1}^s u_r y_{rj} &\leq 1, j = 1, \dots, n \\ u_r &\geq 0, r = 1, \dots, s \end{aligned} \quad (1)$$

Model (1) is computed separately for each firm and the subscript j_0 refers to the firm whose relative performance is under evaluation. y_{rj} is the value of the performance dimension r ($r = 1, \dots, s$) observed for firm j . u_r is the set of weights assigned to performance dimension r to estimate the performance of j_0 in the best possible light.

Model (1) involves finding optimal values for the decision variables u_r . The aggregate performance (corresponding to a composite indicator resulting from the aggregation of the four performance dimensions) for the firm j_0 is maximized, subject to the constraint that the performance measures must be less than or equal to one for all firms in the sample when evaluated with similar weights. If using the optimal weights for firm j_0 no other firm reaches a score of aggregate performance higher than the value assigned to firm j_0 , it implies that the firm defines the frontier. In this case, the objective function of (1) returns a score equal to one. Otherwise, firm j_0 is considered an underperformer, meaning that it is located below the performance frontier.

The estimated performance frontier should correspond to replicable firm-level performance and be robust, that is, it should be well populated and not too distant from the bulk of firms in the analysis (Wilson 1995). In this context, it is important to examine the presence of outlier firms since they may lead to the shift of the frontier and have a significant impact on the evaluation of other firms. There are many methods available in the literature for identifying and handling outliers (Wilson 1995, Ondrich

Table 5 Performance evaluation of firms considering the performance frontier estimated with DEA

Group of firms	Performance dimensions	Min	Max	Mean	St. Dev
Frontier ($n = 22$; 9.2% of the usable sample)	Cost	1.00	5.00	3.02	0.957
	Delivery	2.25	5.00	3.59	0.714
	Flexibility	3.00	5.00	3.99	0.701
	Quality	3.50	5.00	4.36	0.640
	Performance score	100%	100%	100%	0
Below the frontier ($n = 216$; 90.8% of the usable sample)	Cost	1.00	4.50	2.95	0.538
	Delivery	1.00	4.50	3.21	0.541
	Flexibility	1.00	4.50	3.29	0.545
	Quality	1.00	4.50	3.40	0.579
	Performance score	24.71%	98.84%	82.02%	9.38
All firms	Cost	1.00	5.00	2.96	0.584
	Delivery	1.00	5.00	3.24	0.568
	Flexibility	1.00	5.00	3.36	0.590
	Quality	1.00	5.00	3.49	0.660
	Performance score	24.71%	100.00%	83.68%	10.35

and Ruggiero 2002, Johnson and McGinnis 2008, or Boyd et al. 2016). As noted by the recent paper of Boyd et al. (2016), despite the many existing methods, the rules for outlier detection and removal are still ad hoc, since the thresholds defined for exclusion of outliers may imply that one removes true frontier points or that we do not remove all outliers given the existence of close neighbors to outliers. As a result, we decided to perform a simple inspection to our data before any inspection for outliers: we plotted the distribution of the average performance of firms as constituted by a crude measure of the average of QP, CP, FP, and DP. The boxplot of the average performance did not show any outlier firms, which was understood as an indication that outliers were not a serious problem in our sample. As a result, we proceeded to estimate the frontier using model (1) and confirmed that the resulting performance frontier was well populated (9.2% of the firms were located on the frontier).

In the previous section we discussed the issue of contextual (exogenous) factors and the way they influenced our choice of the sample of units to analyze. There are alternatives to handle contextual factors: in particular they could be handled together with model (1). The models that can measure performance while taking into account the context of the unit, are known as conditional BoD or DEA models and are discussed extensively in the literature (see e.g. De Witte and Schiltz 2018, Verschelde and Rogge 2012). In this paper, we opt to employing the generic BoD model, since the estimation of the production frontier is just a means to an end—that is, we are not interested in the computation of robust efficiency scores but mostly on the relationship of these scores with other variables.

BoD models' (such as (1)) optimal solutions may give rise to extreme or unrealistic weights (e.g., zero weights). This has resulted in models that prevent the occurrence of zero weights through weight restrictions, or non-oriented models that force projections on the Pareto-efficient frontier (see Portela and Thanassoulis 2006, for further discussion on the equivalence between these formulations). In this paper we opted for not dealing with zero weights because our objective was the identification of the frontier and its composed units, rather than the performance of each unit per se. In any case, we verified how many frontier units had zero weights, and the results indicated that 8 out of 22 efficient units have 3 zero weights in their assessment, the remaining 14 units have 2, 1, or 0 occurrences of zero weights—and the non-occurrence of zero weights was the rarest. Therefore, we concluded that the frontier indeed has many units with null-weights, meaning that projections to the frontier may happen in a weakly efficient part of the frontier. To check whether the existence of weakly efficient units significantly affected our results, we computed an enhanced Russell measure in the spirit of Pastor et al. (1999) using ratios of pareto-efficient targets and observed values. Correlation between the Russell and the radial measure of efficiency were on the order of 0.97. In addition, only 4 units that had originally a radial efficiency score above 90% obtained a Russell score below the 90% threshold, meaning that they were no longer considered best performers using the non-radial measure. This shows that the main conclusions reached in this paper are not sensitive to the use of a weakly efficient frontier instead of a strong efficient one.

Table 6 Comparison between best performers and underperformers

Performance threshold	Sample split	Average BPP	Size (no. of employees)	GNI/Cap (USD)	%ind25	%ind28
100% ^a	Best performers (9.2%)	3.20	181.05	43,985.5	10.1%	8.0%
	Underperformers	2.83	160.88	41,220.7	89.9%	92.0%
	<i>p</i> value	0.024	0.426	0.506	0.654	
95%	Best performers (17.2%)	3.21	159.56	43,341.9	16.8%	13.0%
	Underperformers	2.78	168.70	41,087.9	83.2%	87.0%
	<i>p</i> value	0.00	0.650	0.286	0.401	
90%	Best performers (30.7%)	3.16	158.55	43,393.8	29.7%	32.0%
	Underperformers	2.73	164.60	40,627.9	70.3%	68.0%
	<i>p</i> value	0.000	0.704	0.109	0.776	

^aThe 100% threshold corresponds to considering as best performers the firms located on the performance frontier

A summary of performance results is provided in Table 5. On average, firms exhibited a performance score of 83.68%. The frontier firms have a score of 100% and show, on average, better levels of performance across all four performance dimensions. Consistent with expectations, these firms exhibit a diverse range of performance mixes across the four dimensions, with minimum values ranging from 1.00 to 3.50. The firms below the frontier exhibit an average performance score of 82.02% meaning that there is a reasonable average potential for improvement identified in our sample of firms (in the order of 18%). The lowest performing firm has a large potential for improvement since its performance score is 24.71%.

Table 6 compares the best performers (firms on or near the frontier) with the underperformers (firms far from the frontier), considering three alternative performance thresholds. It shows that best performing firms have significantly higher levels of BPs (BPP) than underperforming firms, while there are no differences between these two groups in terms of contextual variables (size, GNI per capita of the country where the firm is located, and industry sector).

5.3 The level of implementation of BPs is negatively associated with distance from the performance frontier (H1)

To test H1, we regressed the natural logarithm of aggregate performance as provided by Eq. (1) on BPP, controlling for the relevant contextual variables, namely, firm size (natural logarithm of the number of employees), industry sector (0–1 dummy, where 1 corresponds to sector 25, and 0 to sector 28), and GNI/capita. We employed Ordinary Least Squares because this method provides consistent estimates of the impact of the determinants of performance scores when the scores are expressed in logarithms (Banker and Natarajan 2008). Table 7 presents the regression results for a model with controls only and a model that includes controls and BPP. The

Table 7 Regression analysis on the logarithm of the aggregate performance score

Independent variables	Regression coefficients	
	Controls only	All variables
Ln Size (no. of employees)	0.003 (0.424)	−0.010 (0.789)
GNI/cap	1.07E-6 (0.377)	9.29E-7 (0.489)
Ind25	0.018 (0.145)	0.006 (0.349)
BPP		0.076* (0.000)
<i>R</i> ²	0.012	0.160
Adj. <i>R</i> ²	0.000	0.146
<i>F</i>	0.969 (0.408)	11.114 (0.000)

Significance levels: **p* < 0.001
(*p* values in brackets)

model with controls only is not statistically significant and has a very limited explanatory power. The model that includes BPP is statistically significant and has relevant explanatory power (about 16% of the variation in performance scores can be explained by differences in BPP across firms), showing that BPP has a positive significant impact on aggregate performance (or, equivalently, a significant negative impact on distance to the performance frontier). Thus, H1 is supported.

5.4 Increments in the level of implementation of BPs over time are positively associated with the degree of simultaneous improvements in performance dimensions and the impact of these increments is positively moderated by the distance to the performance frontier (H2 and H3)

To test H2 and H3, we drew on the respondents' assessment of changes in BP implementation (BPC) and performance dimensions (QC, DC, FC, CC) over the previous 3 years. There were no missing values for these variables and therefore no need to exclude further cases. Following the logic in Lapré

and Scudder (2004), we defined the variable Improvement Breadth (IB) as the number of performance dimensions that had improved in the last 3 years. Each of the performance change variables was composed of two items rated on a 1–5 scale, where 1 represented decline, 2 represented no change and 3, 4, and 5 represented different magnitudes of increments (Table 3 and Appendix A). Thus, we considered that each of the performance change variables corresponded to a performance enhancement if its value (aggregate of the two associated items) were above 2.00. As such, IB is an ordinal variable that can take the values of 0, 1, 2, 3 and 4 (0 meaning no dimension improved in the last 3 years, 1 meaning that one dimension improved and so forth). We employed cumulative logit regression to test H2 and H3, implemented via the ordinal regression procedure in SPSS. The ordinal dependent variable was IB, while the core independent variables were BPC (H2), the firm's performance score at time t (present) (PS) given by the DEA analysis (Table 5) and the interaction term BPC*PS (H3). Given the long-term nature of performance potential, we expect that the frontier for the studied group of firms estimated at time t is a good approximation of the frontier at time $t-3$ years. This is because moving the performance frontier of a group of firms normally requires large capital investments and radical changes to the physical plants (Schmenner and Swink 1998) so that this frontier is relatively stable over time when compared to changes in operating practices (Schmenner and Swink 1998; Vastag 2000). We also controlled for firm size, industry sector and GNI per capita.

As shown in Eq. (2), the cumulative logit model estimates a single slope for each independent variable, but different intercepts (β_j) for each category j of the dependent variable (i.e., it assumes a set of parallel lines, one for each level of the dependent variable). The interaction term allows us to test the moderating effect of the distance to the frontier (or equivalently, PS).

The odds term in (2) $\left(\frac{P(IB \leq j)}{1 - P(IB \leq j)}\right)$ compares the probability that IB is equal to or below j with the probability that it is higher than j . Thus, the lower the log-odds term in (2), the more likely it is for IB to take values higher than j . Slopes $\gamma/\alpha/\theta$ are interpreted as the decrease in the log-odds of a firm falling into any category j or below when BPC/PS/BPC*PS increase one unit (Ananth and Kleinbaum 1997) (note that the sign associated with the slopes is negative). Hence, a positive slope indicates that increasing BPC/PS/ BPC*PS tends to increase the probability of the responses falling into higher IB categories.

$$\log\left(\frac{P(IB \leq j)}{1 - P(IB \leq j)}\right) = \beta_j - \gamma BPC - \alpha PS - \partial \text{CONTROLS} - \theta BPC * PS \quad j = 0, \dots, 4 \quad (2)$$

Table 8 Results from the ordinal regression

Independent variables	Coefficient	p value
Ln Size (no. of employees)	0.035	0.841
GNI/cap	1.68E-05	0.108
Ind25	0.133	0.604
Slope α (PS)	0.035*	0.014
Slope γ (BPC)	0.609*	0.001
Slope θ (BPC*PS)	-0.037*	0.023
Quality of regression		
Pseudo R^2 Nagelkerke	18.8%	
Goodness-of-fit tests		
Pearson	870.09	0.954
Deviance	547.47	1.000

Significance level: * $p < 0.05$

A hierarchical procedure was followed using in the first step only controls, adding in the second step BPC and PS and adding in a third step the interaction term BPC*PS. Variables were centered in all cases to avoid multicollinearity problems (Jaccard et al. 1990). The SPSS χ^2 test of parallel lines was performed in all steps, showing that the assumption of equal slopes was not rejected ($\chi^2 = 9.137$ and $p = 0.956$ in the complete model—step 3). As a result, the standard ordinal regression was performed.

In step 1, the model was not statistically significant exhibiting a pseudo R^2 (Nagelkerke 1991) of 2.2%. In step 2, the model become statistically significant and the Pseudo R^2 increased to 16.4%. Table 8 shows the estimation results of the ordinal regression in the final step. The pseudo R^2 is 18.8% and the model fitting χ^2 (of Pearson and Deviance measures) show that there is a good fit (goodness-of-fit tests not rejected). The inclusion of the interaction term increased the variance explained by about 2.4%.

The statistically significant positive γ slope in Table 8 indicates that firms with higher levels of BPC tend to have a higher probability of larger IBs (i.e. more dimensions of performance being improved). Thus, H2 is supported. Similarly, the statistically significant positive α slope means that, as expected, higher performing firms (those closer to the frontier) are more likely to be included into higher IB categories. Finally, statistically significant negative θ slope (interaction term) means that the impact of BPC is negatively moderated by aggregate firm performance or equivalently, the closer the unit gets to the frontier the lower the impact of BPC on improvement breath. Thus, H3 is supported.

5.5 Firms located on or near the performance frontier exhibit stronger trade-offs between performance dimensions (H4)

To test H4, we employed the pervasiveness of positive and negative pairwise correlations across the four performance

Table 9 Pairwise correlations across performance dimensions for best performers and underperformers

Performance dimensions	Underperformers ($n = 198$)			Best performers ($n = 41$)			No. of negative correlations in best performers/underperformers	No. of positive correlations in best performers/underperformers
	(Performance <95%)			(Performance \geq 95%)				
	Cost	Delivery	Flexibility	Cost	Delivery	Flexibility		
							2/0	0/4
Delivery	0.392**			0.299				
Flexibility	0.089	0.381**		-0.231	-0.147			
Quality	0.106	0.399**	0.338**	-0.477**	-0.107	-0.356*		

Significance levels: * $p < 0.05$; ** $p < 0.01$

dimensions (QP, DP, FP, CP) to infer about the extent to which trade-offs occurred for best performers (firms on or near the frontier) and for underperformers (firms far from the frontier). We assessed whether firms were on or near the frontier based on the 95% aggregate performance threshold. This threshold balances the need for an adequate sample size of best performers with the need to discriminate between these firms and underperformers.

Table 9 shows the pairwise correlations across the four performance dimensions. All significant correlations in the underperformer group are positive, suggesting that the competitive repositioning by the increase in one performance dimension may be accomplished with the simultaneous increase in other performance dimensions. For the best performers group, one or more significant correlations are negative (as expected, at least for the frontier units, given the inherent DEA assumptions). The number of positive correlations is much higher for the underperformers, than for the best performers. Taken together, the results lend support for H4. An inspection of the values of the four performance dimensions across firms revealed that no single firm exhibited superior performance in all dimensions simultaneously (strong-form domination), which reinforces the existence of repositioning trade-offs among best performers. Interestingly, the trade-offs revealed are essentially between cost and quality and also quality and flexibility.

As we saw in Table 6, the comparisons between best performers and underperformers are robust to the chosen performance threshold. Despite that, we performed a robustness check on the threshold value used to distinguish between the two groups of firms, considering alternative thresholds of 85% and 90% (Table 10). For each threshold, we found no negative significant correlations in underperformer firms. For all thresholds, we also found a higher number of positive correlations in underperformers than in best performers. This shows that the results are robust irrespective of the chosen threshold.

5.6 Robustness checks

We assessed the robustness of our findings in relation to industry and firm size. To do so, we repeated the analyses conducted for the baseline group (G0; $n = 238$) for two other groups of firms within high-income countries. The first group (G1) comprised SMEs from the union of industry sectors ISIC 26 and 27 ($n = 70$). These two sectors comprise the manufacture of electronic, electric, and optical products. The second group (G2) comprised large firms (number of employees above 500) from the union of industry sectors ISIC 25 and 28 ($n = 95$). Appendix B presents the results of the analyses. The results show that H1 and H2 are supported in G1 and G2 ($p < 0.05$). H3 is supported in G1, but only at $p = 0.10$, and not supported in G2. H4 receives only limited support in G1 and good support in G2. The decreased support for H3 and H4 found in G1 and G2, when compared to G0, is coherent with the reduced samples sizes (decreasing the power of the statistical tests) and reduced variation in performance scores (decreasing the ability to capture the impact of distance to the frontier). Overall, the findings for G1 and G2 are consistent with those found in the baseline group.

6 Theoretical and managerial implications

Our findings have a number of implications for theory. The results indicate that BPs can support a diverse range of dominant, difficult-to-imitate competitive positions (H1). The associated advantage is in relation not only to actual peers, but also to notional firms that might enter the market. This is a novel finding, which extends results from correlation studies of positive associations between BPs and individual performance dimensions (da Silveira and Sousa 2010; Shah and Ward 2003). Our study is one of the first to empirically establish a clear link between the use of BPs and distinctive operations-based capabilities in manufacturing. From a conceptual perspective, this finding supports the view of BPs as sources of enduring performance

Table 10 Robustness checks on H4 testing

Performance score threshold	% of best performers	No. of negative correlations in best performers/underperformers	No. of positive correlations in best performers/underperformers
85%	43.93	2/0	1/4
90%	30.96	2/0	1/4

The counts of positive and negative correlations refer to $p < 0.05$

advantage, reconciling the resource-based theory and practice-based view perspectives (Hitt et al. 2016).

Over time, more intense BP implementations may lead to simultaneous improvements on multiple performance dimensions (H2). However, this effect appears to decrease as the firm gets closer to the performance frontier (H3), leading to more trade-offs between performance dimensions (H4). The study unveils at least two important aspects of trade-off management in operations.

The first, is that trade-offs may constrain competitiveness even in a multidimensional performance context. Previous studies, e.g., Mapes et al. (1997) and Lapre and Scudder (2004) focused on bi-variate correlations between performance dyads, which led to limited insight on the nature and strength of trade-offs. Our analysis indicates that trade-offs may still diminish BP returns, even when the performance frontier is defined by multiple outputs. The second, refers to the validity of trade-offs in operations. Our findings support the view that underperforming firms operate more often under the law of cumulative capabilities, while best performing firms are more frequently bound by the law of trade-offs (Schmenner and Swink 1998). The use of frontier analysis provides a different interpretation for the much stronger support received by the law of cumulative capabilities in prior empirical research. In fact, as conjectured by Rosenzweig and Easton's (2010) meta-analysis, such support may result from trade-offs being empirically masked by performance enhancements that occur below the performance frontier. This argument was proposed by authors such as da Silveira and Slack (2001), but often missed by studies "counting" the number of trade-offs in business units. Consistent with this, we found that in our sample no single firm exhibited superior performance in all dimensions simultaneously (strong-form domination). Our results suggest that distinctive positions within best performing firms may be strongly determined by hard manufacturing strategic choices, subject to trade-offs. This is an important contribution to the long-standing debate on whether and when manufacturing trade-offs are real (Rosenzweig and Easton 2010).

Our study informs the efficiency and productivity analysis literature by employing efficiency measurement methods in an operations context. This has been an under-researched area (exceptions include Brea-Solís et al. 2015; Triebis and Kumbhakar 2018; and Epure 2016), and, to the authors' knowledge, there has not been the application of the benefit of the doubt approach to compute an aggregate index of operational

performance. Indeed, Brea-Solís et al. (2015), Triebis and Kumbhakar (2018) and Epure (2016) all link operations aspects of the firm to the traditional concept of efficiency—the process through which inputs like labor and capital are used to produce outputs like revenue—and do not compute aggregate performance as we do in this paper. So, the "efficiency" concept implicit in our analysis is an operations concept computed through efficiency analysis methods.

The study also informs managerial practice. The results indicate that firms can leverage BPs to reach various dominant and difficult-to-imitate competitive positions. This dynamic nature of BPs can be especially useful in turbulent and uncertain markets. The study may also help managers to sustain established BP programs over time, which is an important concern as programs mature (Hines et al. 2011). By providing evidence of diminishing returns, the study highlights the need to adjust expectations about BP programs as the firm approaches higher levels of performance and program maturity (Netland and Ferdows 2016). At this stage, BPs may be important to maintain (defend) advantageous competitive positions, in two ways. First, they may allow a best performing firm to remain on the frontier if it decides to re-position across the competitive space. Second, they may allow to close gaps that may emerge if the performance frontier of its industry shifts out, e.g., due to the introduction of disruptive technologies. Besides, firms need to explicitly consider the effects of trade-offs as they move along improvement trajectories and change strategic positions. This understanding may help towards behavioral adjustment and leadership actions (Tortorella et al. 2018) to sustain motivation and internal support for BPs use. Third, the study highlights dynamic aspects of BPs implementation, such as organizational learning and program adaptation, as important sources of advantage. Especially as firms get closer to their industry's performance frontier, they can benefit from carefully tailoring and adapting the mix of BPs to their context and competitive position, rather than just implementing BPs by imitation. Thus, managers should pay attention to creating an environment in the manufacturing organization that promotes these aspects in BP implementation.

The results also indicate that managers must take an all-inclusive approach incorporating multiple outputs for the analysis of trade-offs in operations. Implementing BPs may lead not only to multiple improvements in the short term, but also (and perhaps paradoxically) stronger trade-offs in the long term, which diminishes rather than increases the returns of

future implementations. This serves as reminder that even the more “local” improvement plans should be informed by operations strategies addressing the complex mix of capabilities, practices, and performance goals of the organization. No matter how small, every decision about operations structure or infrastructure may affect multiple performance areas. For example, it has been recently reported (Engle 2021) that Tesla’s challenges with NPD might have simultaneous effects on the flexibility (mix), quality, and cost competitiveness of their products. Thus, improving on Tesla’s NPD program might have potential effects on various dimensions, bringing the company ever closer to the industry performance frontier but also uncovering trade-offs that might not be currently evident in the organization.

7 Conclusion

Our study contributes to manufacturing strategy and productivity analysis research in several ways. First, it shows how researchers can adopt the theoretical and empirical underpinnings of frontier analysis to explore issues of performance advantage and operations-based capabilities. Theoretically, we explain how frontier analysis can integrate the resource-based theory and practice-based view perspectives of BPs in operations, through the concept of dynamic capabilities. The incorporation of these perspectives in DEA-related studies has been seldom done, while some efforts exist (e.g., Brea-Solís et al. 2015; Epure 2016). Empirically, we show how frontier analysis can be used to capture performance advantage and operations-based capabilities by estimating distance to the performance frontier.

Second, with the benefit of this approach, we carried out a rigorous analysis of key elements of manufacturing strategy theory. Specific advances in research design relative to prior studies include: (i) estimation of performance frontiers based on a comprehensive set of operational performance dimensions; (ii) estimation of diminishing returns and trade-offs based on distance to performance frontiers; (iii) use of longitudinal data (even if based on retrospection) to examine trade-offs.

Third, the study provides novel insights about the role of BPs as sources of enduring performance advantage. We articulate the notion of BP implementations as dynamic capabilities based on their ability to (i) lead to a heterogeneous range of dominant competitive positions through improvements in multiple performance dimensions, idiosyncratic integrations between practices and redirection of performance objectives of individual practices over time and (ii) generate difficult-to-imitate positions via second-order learning and unique improvement trajectories.

Fourth, the study shows that trade-offs become stronger the closer the unit moves to the frontier, contributing to the

long-standing debate on whether and when manufacturing trade-offs occur (Rosenzweig and Easton 2010).

Our study suggests relevant avenues for future research on BPs implementation. Recent studies focused on behavioral aspects of implementation including leadership, employee behavior and sustained involvement (Fundin et al. 2018, 2019; Tortorella et al. 2018). Our evidence to diminishing returns as firms approach the performance frontier suggests that research should consider distance to the performance frontier as a contingency factor that might affect expectations by managers and employees about program benefits. The notion of BPs implementation as leading to dynamic capabilities draws attention to the need to investigate leadership and motivational strategies that enable organizational learning and dynamic adaptations of BP programs over time (Yang 2008).

While our study overcomes limitations of prior research, there are also opportunities for future studies to adopt research designs that complement ours. We captured longitudinal change based on respondents’ retrospective assessments of changes in BPs and performance over 3 years. Future research could collect panel data to capture adoption and performance trajectories in real time (Rosenzweig and Easton 2010). It could also examine how long a specific practice can sustain performance variation (i.e., the extent to which the resulting advantage is sustained) (Bromiley and Rau 2014). We estimated performance frontiers based on one specific technique (DEA), with specific strengths (e.g., no a priori assumptions regarding the functional form of the frontier), but also some limitations (e.g., dependency on extreme observations and non-statistical nature of the distance scores). Future research could employ techniques with different strengths (like bootstrapping to account for noise in the distance estimates, or partial frontier models, such as order-m, to avoid the dependency on extreme observations or outliers and problems of zero weights). These issues were not addressed in this paper because they were not deemed central to the analysis and because a diagnosis of these problems revealed that they were not serious. However, there is room for employing other methods for frontier estimation and compare those with the one employed in this paper.

Acknowledgements MCS and RS are grateful to the Portuguese Foundation for Science and Technology (FCT) for the support to CEGE - Research Centre in Management and Economics (project UIDB/00731/2020). ASC acknowledges the Portuguese Foundation for Science and Technology (FCT) for financial support via the project eduBEST: Education Systems Benchmarking with Frontier Techniques (2022.08686.PTDC).

Funding Open access funding provided by FCTIFCCN (b-on).

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

8 Appendix A. Measures (Source: IMSS-VI)

Variable	<i>n</i>	μ	σ
<i>Current level of implementation of action programs</i>			
<i>Indicate [...] the current level of implementation of: (None [1], High [5])</i>			
TQM – present (TQMP, $\alpha = 0.84$, $n = 896$)			
Quality improvement and control	911	3.36	1.11
Improving equipment availability	909	3.22	1.12
Benchmarking/self-assessment	902	2.94	1.25
Lean – present (LEANP, $\alpha = 0.74$, $n = 902$)			
Restructuring processes and layout	910	3.40	1.04
Pull production	904	3.27	1.12
NPD – present (NPDP, $\alpha = 0.78$, $n = 871$)			
Design integration	892	3.23	1.06
Organizational integration	892	3.17	1.08
Technological integration	876	3.09	1.18
<i>Current performance compared with main competitor(s)</i>			
<i>How does your current performance compare with that of your main competitor(s)? (Much lower [1], Much higher [5])</i>			
Cost performance (CP, $\alpha = 0.71$, $n = 811$)			
Unit manufacturing cost ^a	835	3.01	0.79
Ordering costs ^a	814	3.04	0.68
Delivery performance (DP, $\alpha = 0.68$, $n = 813$)			
Delivery speed	875	3.55	0.83
Delivery reliability	875	3.58	0.84
Manufacturing lead time ^a	829	3.19	0.78
Procurement lead time ^a	823	3.08	0.74
Flexibility performance (FP, $\alpha = 0.74$, $n = 842$)			
Volume flexibility	863	3.45	0.80
Mix flexibility	863	3.43	0.80
Product customization ability	862	3.52	0.85

Table (continued)

Variable	<i>n</i>	μ	σ
New product introduction ability	869	3.45	0.93
Quality performance (QP, $\alpha = 0.79$, $n = 877$)			
Conformance quality	880	3.51	0.76
Product quality and reliability	881	3.65	0.80
<i>Effort in the last 3 years into implementing programs:</i>			
<i>Indicate the effort put in the last 3 years into implementing: (None [1], High [5])</i>			
TQM – effort (TQMC, $\alpha = 0.84$, $n = 898$)			
Quality improvement and control	908	3.28	1.10
Improving equipment availability	905	3.12	1.11
Benchmarking/self-assessment	900	2.87	1.23
Lean – effort (LEANC, $\alpha = 0.74$, $n = 893$)			
Restructuring processes and layout	899	3.42	1.12
Pull production	896	3.23	1.18
NPD – effort (NPDC, $\alpha = 0.79$, $n = 874$)			
Design integration	891	3.17	1.09
Organizational integration	888	3.05	1.07
Technological integration	878	3.02	1.18
<i>Manufacturing performance change over the last 3 years</i>			
<i>How has your manufacturing performance changed over the last 3 years? (Decreased (5% or worse) [1], Stayed about the same (-5%/+5%) [2], Slightly increased (+5–15%) [3], Increased (+15–25%) [4], Strongly increased (+25% or better) [5])</i>			
Cost change (CC, $\alpha = 0.75$, $n = 878$)			
Unit manufacturing cost ^a	893	2.51	0.97
Ordering costs ^a	881	2.42	0.88
Delivery change (DC, $\alpha = 0.76$, $n = 877$)			
Delivery speed	909	3.20	0.97
Delivery reliability	908	3.25	0.99
Manufacturing lead time ^a	895	2.81	0.97
Procurement lead time ^a	890	2.62	0.90
Flexibility change (FC, $\alpha = 0.77$, $n = 886$)			
Volume flexibility	907	3.24	1.00
Mix flexibility	896	3.15	0.96
Product customization ability	904	3.07	1.01
New product introduction ability	904	3.21	1.02
Quality change (QC, $\alpha = 0.84$, $n = 908$)			
Conformance quality	910	3.13	0.94
Product quality and reliability	911	3.26	0.97

^aReversed scales

Table 11 Statistics on the performance scores

Group	Average performance score	St. Dev	Min	Max	No. (%) of firms on the frontier
G1 ($n = 70$)	87.50%	10.47	54.03%	100%	18 (24.7%)
G2 ($n = 95$)	85.53%	8.98	64.44%	100%	12 (12.6%)

Table 12 Statistics on the determinants of performance for firms on the frontier and firms below the frontier

Group	Sample split	Average BPP	Size (no. of employees)	GNI/Cap ^a (USD)	%ind25 or %ind26	%ind28 or %ind27
G1	Frontier (24.7%)	3.42	239.67	36,206.6	19.05%	27.5%
	Below the frontier	3.06	172.43	40,901.3	80.95%	72.5%
	<i>p</i> value	0.063	0.063	0.093	0.558	
G2	Frontier (12.6%)	3.49	2800.17	42,472.3	18.4%	6.5%
	Below the frontier	3.21	6920.25	40,114.34	81.6%	93.5%
	<i>p</i> value	0.144	0.261	0.506	0.122	

^aSource: World Bank (2019), data for 2012

Table 13 Regression analysis on the logarithm of aggregate performance score

Independent variables	Regression coefficients	
	G1	G2
Ln Size (no. of employees)	0.024	-0.007
GNI/cap	2.9E-07	-1.11E-6
Ind25 OR Ind26	-0.013	0.000
BPP	0.058**	0.036*
R^2	15.6%	7.6%
Adj. R^2	10.5%	3.4%
F (<i>p</i> value)	3.057(0.023)	1.83 (0.130)

Significance levels: * $p < 0.05$; ** $p < 0.01$

9 Appendix B. Robustness checks on hypotheses tests

The analyses reported in the main text were conducted for a baseline group of SME firms from the union of industry sectors ISIC 25 and 28 in high-income countries (group G0). This Appendix summarizes the analyses for two additional groups of comparable firms in the IMSS sample with feasible, albeit reduced, sample sizes: (i) G1: SME firms in high-income countries from the union of industry sectors ISIC 26 and 27 ($n = 82$); (ii) G2: Large firms in high-income countries from the union of industry sectors ISIC 25 and 28 ($n = 115$).

We addressed missing data on the four first-order variables involved in the estimation of performance frontiers (QP, DP, FP, and CP) for the two groups. Little's (1988) test on these variables was non-significant at the 5% level (for G1: $\chi^2 = 5.182$; $DF = 5$, $p = 0.394$ and for G2: $\chi^2 = 9.742$;

$DF = 10$, $p = 0.463$). Therefore, data could be considered missing completely at random (MCAR) and listwise deletion of cases with missing data was conducted. This resulted in usable samples of 70 cases for G1 and 95 cases for G2. An outlier analysis was also performed and there were no reasons for excluding any firms due to outlier behavior on the four performance variables in these two groups.

9.1 Characterization of groups

Tables 11 and 12 characterize groups G1 and G2. Compared to G0 ($n = 238$), G1 and G2 have much lower sizes ($n = 70$ and $n = 95$, respectively). In addition, G1 and G2 have a narrower range and variation of performance scores, especially G2 (Table 11 vs Table 5). G2 has a less-well populated performance frontier (12 firms), when compared to G0 (22) and G1 (18) (Table 11 vs Table 5). As in G0, in G1 and G2 there are no statistically significant ($p < 0.05$) differences in size, GNI/capita and industry sector between firms on the performance frontier and firms below the performance frontier (Table 12). Given the lower samples sizes of G1 and G2, in the ensuing analyses we report $p = 0.05$ and $p = 0.10$ significance levels, when appropriate.

9.2 Testing of H1

Table 13 shows that, in groups G1 and G2, BPP has a statistically significant positive impact on the aggregate performance score, with controls being non-significant. Thus, H1 is supported in both groups.

9.3 Testing of H2 and H3

Table 14 summarizes the ordinal regression results for groups G1 and G2. The following results were similar in

both groups. The SPSS χ^2 test of parallel lines showed that the assumption of equal slopes was not rejected. The quality of the regression was confirmed by the non-rejection of Pearson and deviance goodness-of-fit tests. The positive

performance score effect is positive and significant at $p = 0.10$, but not at $p = 0.05$. BPC has a statistically significant ($p = 0.05$) positive impact on IB, thus supporting H2.

The moderation effect BPC*PS is negative and significant in G1 at $p = 0.10$, but not at $p = 0.05$. This effect is not significant in G2. Thus, H3 is supported in G1 at a lower level of significance ($p = 0.10$) and is not supported in G2.

Table 14 Results from the ordinal regressions

Independent variables	G1		G2	
	Coefficient	<i>p</i> value	Coefficient	<i>p</i> value
Ln Size (no. of employees)	0.177	0.479	0.103	0.543
GNI/cap	-3.24E-6	0.888	-2.77E-5	0.186
Ind25 OR Ind26	-0.830	0.098	0.061	0.889
Slope α (PS)	0.048 [†]	0.056	0.047 [†]	0.065
Slope γ (BPC)	0.798*	0.028	1.091***	0.001
Slope θ (BPC*PS)	-0.058 [†]	0.097	0.003	0.930
Quality of regression				
Pseudo R^2	18.50%		28.00%	
Nagelkerke				
Goodness of fit tests				
Pearson	264.20	0.588	280.45	0.999
Deviance	185.16	1.000	191.32	1.000

Significance levels: [†] $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Significant parameters in bold ($p < 0.05$)

9.4 Testing of H4

Tables 15 and 16 summarize the results. H4 receives only limited support in G1, in the form of a higher number of positive correlations for the underperforming firms for the specific performance threshold of 85% and one more negative correlation ($p = 0.10$) for best performers for the specific performance threshold of 95%. H4 receives good support in G2 in the form of higher/lower number of negative/positive ($p = 0.05$) correlations for best performers/underperformers across several performance thresholds. We also found that no firm strong-form dominated other firms in G1 and G2.

Table 15 Pairwise correlations across performance dimensions for best performers and underperformers

Group: no. of underperformers/no. of best performers/% of best performers	Best performers (performance $\geq 95\%$)			Underperformers (performance $< 95\%$)			No. of negative correlations in best performers/underperformers	No. of positive correlations in best performers/underperformers
	Cost	Delivery	Flexibility	Cost	Delivery	Flexibility		
	G1: 47/23/32.8%							
Delivery	0.512*			0.323*				
Flexibility	-0.288	-0.082		0.056	0.078			
Quality	-0.022	0.238	-0.384 [†]	-0.093	0.035	0.212		
G2: 79/16/16.8%								
Delivery	0.390			0.077				
Flexibility	-0.613*	-0.322		0.006	0.304**		1/0	0/3
Quality	-0.285	0.457 [†]	-0.078	-0.112	0.377**	0.326**		

Significance levels: [†] $p < 0.10$; * $p < 0.05$; ** $p < 0.01$. Significant parameters in bold ($p < 0.05$)

Table 16 Robustness checks on H4 testing

Performance score threshold	% of best performers	No. of negative correlations in best performers/underperformers	No. of positive correlations in best performers/underperformers
85% G1	56.94	0/0	1/2
G2	51.60	0/0	1/2
90% G1	40.28	0/0	1/1
G2	32.60	0/0	0/3

The counts of positive and negative correlations refer to $p < 0.05$. The counts are the same for $p < 0.10$

References

- Ahmed N, Montagnò R, Firenze R (1996) Operations strategy and organizational performance: an empirical study. *Int J Oper Prod Manag* 16(5):41–53. <https://doi.org/10.1108/01443579610113933>
- Anand G, Ward P, Tatikonda M, Schilling D (2009) Dynamic capabilities through continuous improvement infrastructure. *J Oper Manag* 27(6):444–461. <https://doi.org/10.1016/j.jom.2009.02.002>
- Ananth C, Kleinbaum D (1997) Regression models for ordinal responses: a review of methods and applications. *Int J Epidemiol* 26(6):1323–1333. <https://doi.org/10.1093/ije/26.6.1323>
- Avella L, Vazquez-Bustelo D, Fernandez E (2011) Cumulative manufacturing capabilities: an extended model and new empirical evidence. *Int J Prod Res* 49(3):707–729. <https://doi.org/10.1080/00207540903460224>
- Bagozzi R, Yi Y (1988) On the evaluation of structural equation models. *J Acad Mark Sci* 16(1):74–94. <https://doi.org/10.1007/BF02723327>
- Banker R, Natarajan R (2008) Evaluating contextual variables affecting productivity using data envelopment analysis. *Oper Res* 56(1):48–58. <https://doi.org/10.1007/BF02723327>
- Bititci U, Ackermann F, Ates A, Davies J, Gibb S, MacBryde J, Mackay D, Maguire C, Van der Meer R, Shafti F (2011) Managerial processes: an operations management perspective towards dynamic capabilities. *Prod Plan Control* 22(2):157–173. <https://doi.org/10.1080/09537281003738860>
- Bevilacqua M, Ciarapica F, de Sanctis I (2017) Lean practices implementation and their relationships with operational responsiveness and company performance: an Italian study. *Int J Prod Res* 55(3):769–794. <https://doi.org/10.1080/00207543.2016.1211346>
- Boyd T, Docken G, Ruggiero J (2016) Outliers in data envelopment analysis. *J Cent Cathedra* 9(2):168–183. <https://doi.org/10.1108/JCC-09-2016-0010>
- Brea-Solís H, Casadesus-Masanell R, Grifell-Tatjé E (2015) Business model evaluation: quantifying walmart's sources of advantage. *Strateg Entrepreneurship J* 9(1):12–33. <https://doi.org/10.1002/sej.1190>
- Bromiley P, Rau D (2014) Towards a practice-based view of strategy. *Strateg Manag J* 35(8):1249–1256. <https://doi.org/10.1002/smj.2238>
- Bromiley P, Rau D (2016) Operations management and the resource-based view: another view. *J Oper Manag* 41(1):95–106. <https://doi.org/10.1016/j.jom.2015.11.003>
- Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur J Oper Res* 2(6):429–444. [https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8)
- Chatha K, Butt I (2015) Themes of study in manufacturing strategy literature. *Int J Oper Prod Manag* 35(4):604–698. <https://doi.org/10.1108/IJOPM-07-2013-0328>
- Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T (2007) An introduction to 'benefit of the doubt' composite indicators. *Soc Indic Res* 82(1):111–145. <https://doi.org/10.1007/s11205-006-9029-7>
- Cooper W, Seiford L, Zhu J (2011) Handbook on data envelopment analysis. Vol. 164. Springer Science & Business Media
- Dabhilkar M, Bengtsson L (2008) Invest or divest? On the relative improvement potential in outsourcing manufacturing. *Prod Plan Control* 19(3):212–228. <https://doi.org/10.1080/09537280701830144>
- da Silveira G, Sousa R (2010) Paradigms of choice in manufacturing strategy: exploring performance relationships of fit, BPs, and capability-based approaches. *Int J Oper Prod Manag* 30(12):1219–1245. <https://doi.org/10.1108/01443571011094244>
- da Silveira G, Slack N (2001) Exploring the trade-off concept. *Int J Oper Prod Manag* 21(7):949–964. <https://doi.org/10.1108/01443570110393432>
- De Witte K, Schiltz F (2018) Measuring and explaining organizational effectiveness of school districts: evidence from a robust and conditional Benefit-of-the-Doubt approach. *Eur J Oper Res* 267(3):1172–1181
- Devinney T, Yip G, Johnson G (2010) Using frontier analysis to evaluate company performance. *Br J Manag* 21(4):921–938. <https://doi.org/10.1111/j.1467-8551.2009.00650.x>
- Dyson R, Allen R, Camanho A, Podinovski V, Sarrico C, Shale E (2001) Pitfalls and protocols in DEA. *Eur J Oper Res* 132(2):245–259. [https://doi.org/10.1016/S0377-2217\(00\)00149-1](https://doi.org/10.1016/S0377-2217(00)00149-1)
- Engle J (2021) Tesla: weak product pipeline threatens growth narrative, January 10. <https://seekingalpha.com/article/4398129-tesla-weak-product-pipeline-threatens-growth-narrative>. Accessed February 16, 2021
- Epure M (2016) Benchmarking for routines and organizational knowledge: a managerial accounting approach with performance feedback. *J Product Anal* 46:87–107. <https://doi.org/10.1007/s1123-016-0475-1>
- Fichman M, Cummings J (2003) Multiple imputation for missing data: Making the most of what you know. *Organ Res Methods* 6(3):282–308. <https://doi.org/10.1177/1094428103255532>
- Fornell C, Larcker D (1981) Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 18(1):39–50. <https://doi.org/10.2307/3151312>
- Fundin A, Bergquist B, Eriksson H, Gremyr I (2018) Challenges and propositions for research in quality management. *Int J Prod Econ* 199(2018):125–137. <https://doi.org/10.1016/j.ijpe.2018.02.020>
- Furlan A, Galeazzo A, Paggiaro A (2019) Organizational and perceived learning in the workplace: a multilevel perspective on employees' problem solving. *Organ Sci* 30(2):280–297. <https://doi.org/10.1287/orsc.2018.1274>
- Hair J, Anderson R, Babin B, Black W (2010) Multivariate data analysis: a global perspective (Vol. 7). Pearson, Upper Saddle River, NJ
- Helfat C, Winter S (2011) Untangling dynamic and operational capabilities: strategy for the (N) ever-changing world. *Strateg Manag J* 32(11):1243–1250. <https://doi.org/10.1002/smj.955>
- Hines P, Found P, Griffiths G, Harrison R (2011) Staying lean: thriving, not just surviving: CRC Press
- Hitt M, Carnes C, Xu K (2016) A current view of resource-based theory in operations management: a response to Bromiley and Rau. *J Oper Manag* 41(10):107–109. <https://doi.org/10.1016/j.jom.2015.11.004>
- Hitt M, Xu K, Carnes C (2016) Resource-based theory in operations management research. *J Oper Manag* 41(1):77–94. <https://doi.org/10.1016/j.jom.2015.11.002>
- Hjalmarsson L, Kumbhakar S, Heshmati A (1996) DEA, DFA and SFA: a comparison. *J Product Anal* 7(2-3):303–327. <https://doi.org/10.1007/BF00157046>
- International Trade Commission. 2010. Small and medium-sized enterprises: overview of participation in US exports. Investigation No. 332-508 USITC Publication 4125 January. <https://www.usitc.gov/publications/332/pub4125.pdf>
- Jaccard J, Wan C, Turrissi R (1990) The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behav Res* 25(4):467–478. https://doi.org/10.1207/s15327906mbr2504_4
- Johnson A, McGinnis L (2008) Outlier detection in two-stage semi-parametric DEA models. *Eur J Oper Res* 187(2):629–635. <https://doi.org/10.1016/j.ejor.2007.03.041>
- Josse J, Husson F (2012) Selecting the number of components in principal component analysis using cross-validation approximations. *Comput Stat Data Anal* 56(6):1869–1879. <https://doi.org/10.1016/j.csda.2011.11.012>
- Koufteros X, Babbar S, Kaighobadi M (2009) A paradigm for examining second-order factor models employing structural

- equation modeling. *Int J Prod Econ* 120(2):633–652. <https://doi.org/10.1016/j.ijpe.2009.04.010>
- Lapré M, Scudder G (2004) Performance improvement paths in the US airline industry: linking trade-offs to asset frontiers. *Prod Oper Manag* 13(2):123–134. <https://doi.org/10.1111/j.1937-5956.2004.tb00149.x>
- Lieberman M, Dhawan R (2005) Assessing the resource base of Japanese and US auto producers: a stochastic frontier production function approach. *Manage Sci* 51(7):1060–1075. <https://doi.org/10.2307/20110398>
- Linderman K, Schroeder R, Sanders J (2010) A knowledge framework underlying process management. *Decis Sci* 41(4):689–719. <https://doi.org/10.1111/j.1540-5915.2010.00286.x>
- Little R (1988) A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc* 83(404):1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Mackelprang A, Nair A (2010) Relationship between just-in-time manufacturing practices and performance: a meta-analytic investigation. *J Oper Manag* 28(4):283–302. <https://doi.org/10.1016/j.jom.2009.10.002>
- Mapes J, New C, Szejcziwski M (1997) Performance trade-offs in manufacturing plants. *Int J Oper Prod Manag* 17(10):1020–1033. <https://doi.org/10.1108/01443579710177031>
- Nagelkerke N (1991) A note on a general definition of the coefficient of determination. *Biometrika* 78(3):691–692. <https://doi.org/10.1093/biomet/78.3.691>
- Netland T, Ferdows K (2016) The S-curve effect of lean implementation. *Prod Oper Manag* 25(6):1106–1120. <https://doi.org/10.1111/poms.12539>
- OECD (2008) Handbook on constructing composite indicators, methodology and user guide. OECD Publishing
- Ondrich J, Ruggiero J (2002) Outlier detection in data envelopment analysis: an analysis of jackknifing. *J Oper Res Soc* 53(3):342–346. <https://doi.org/10.1057/palgrave/jors/2601290>
- Paashuis V, Boer H (1997) Organizing for concurrent engineering: an integration mechanism framework. *Integr Manuf Syst* 8(2):79–89. <https://doi.org/10.1108/09576069710165765>
- Pastor JT, Ruiz JL, Sirvent I (1999) An enhanced DEA Russell graph efficiency measure. *Eur J Oper Res* 115(3):596–607
- Peng D, Schroeder R, Shah R (2008) Linking routines to operations capabilities: a new perspective. *J Oper Manag* 26(6):730–748. <https://doi.org/10.1016/j.jom.2007.11.001>
- Peters C, Enders C (2002) A primer for the estimation of structural equation models in the presence of missing data: maximum likelihood algorithms. *J Target Meas Anal Mark* 11(1):81–95. <https://doi.org/10.1057/palgrave.jt.5740069>
- Podsakoff P, MacKenzie S, Lee J, Podsakoff N (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol* 88(5):879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Portela MCAS, Thanassoulis E (2006) Zero weights and non-zero slacks: different solutions to the same problem. *Ann Oper Res* 145(1):129–147. <https://doi.org/10.1007/s10479-006-0029-4>
- R-Core-Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Austria: Vienna. <https://www.R-project.org/>. Accessed April 26, 2018
- Rosenzweig E, Easton G (2010) Tradeoffs in manufacturing? A meta-analysis and critique of the literature. *Prod Oper Manag* 19(2):127–141. <https://doi.org/10.1111/j.1937-5956.2009.01072.x>
- Rosseel Y (2012) Lavaan: an R package for structural equation modeling and more. Version 0.5–12 (BETA). *J Stat Softw* 48(2):1–36. <https://doi.org/10.18637/jss.v048.i02>
- RStudio_Team (2015) Integrated Development Environment for R. RStudio, Inc. Boston, USA. www.rstudio.com
- Sarmiento R, Shukla V (2011) Zero-sum and frontier trade-offs: an investigation on compromises and compatibilities amongst manufacturing capabilities. *Int J Prod Res* 49(7):2001–2017. <https://doi.org/10.1080/00207540903555544>
- Schmenner R, Swink M (1998) On theory in operations management. *J Oper Manag* 17(1):97–113. [https://doi.org/10.1016/S0272-6963\(98\)00028-X](https://doi.org/10.1016/S0272-6963(98)00028-X)
- Schonberger R (1996) World class manufacturing. Free Press, New York
- Schonberger R (2019) The disintegration of lean manufacturing and lean management. *Bus Horiz* 62(3):359–371. <https://doi.org/10.1016/j.bushor.2019.01.004>
- Schroeder R, Bates K, Junttila M (2002) A resource-based view of manufacturing strategy and the relationship to manufacturing performance. *Strateg Manag J* 23(2):105–117. <https://doi.org/10.1002/smj.213>
- Schroeder R, Shah R, Peng D (2011) The cumulative capability ‘sand cone’ model revisited: a new perspective for manufacturing strategy. *Int J Prod Res* 49(16):4879–4901. <https://doi.org/10.1080/00207543.2010.509116>
- Shah R, Ward P (2003) Lean manufacturing: context, practice bundles, and performance. *J Oper Manag* 21(2):129–149. [https://doi.org/10.1016/S0272-6963\(02\)00108-0](https://doi.org/10.1016/S0272-6963(02)00108-0)
- Simon D, Hitt M, Ireland R, Gilbert B (2011) Resource orchestration to create competitive advantage: breadth, depth, and life cycle effects. *J Manag* 37(5):1390–1412. <https://doi.org/10.1177/0149206310385695>
- Sousa R, da Silveira G (2019) The relationship between servitization and customization strategies. *Int J Oper Prod Manag* 39(3):454–474. <https://doi.org/10.1108/IJOPM-03-2018-0177>
- Sousa R, Voss C (2002) Quality management re-visited: a reflective review and agenda for future research. *J Oper Manag* 20(1):91–109. [https://doi.org/10.1016/S0272-6963\(01\)00088-2](https://doi.org/10.1016/S0272-6963(01)00088-2)
- Sousa R, Voss C (2008) Contingency research in operations management practices. *J Oper Manag* 26(6):697–713. <https://doi.org/10.1016/j.jom.2008.06.001>
- Spear S (2004) Learning to lead at Toyota. *Harv Bus Rev* 82(5):78–91
- Swink M, Narasimhan R, Kim S (2005) Manufacturing practices and strategy integration: effects on cost efficiency, flexibility, and market-based performance. *Decis Sci* 36(3):427–457. <https://doi.org/10.1111/j.1540-5414.2005.00079.x>
- Tortorella G, Fettermann D, Frank A, Marodin G (2018) Lean manufacturing implementation: leadership styles and contextual variables. *Int J Oper Prod Manag* 38(5):1205–1227. <https://doi.org/10.1108/IJOPM-08-2016-0453>
- Triebts T, Kumbhakar S (2018) Management in production: from unobserved to observed. *J Product Anal* 49:111–1121
- Vastag G (2000) The theory of performance frontiers. *J Oper Manag* 18(3):353–360. [https://doi.org/10.1016/S0272-6963\(99\)00024-8](https://doi.org/10.1016/S0272-6963(99)00024-8)
- Vershelde M, Rogge N (2012) An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: evidence from a conditional Data Envelopment Analysis approach. *Eur J Oper Res* 223(1):214–225
- Voorhees C, Brady M, Calantone R, Ramirez E (2016) Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies. *J Acad Mark Sci* 44(1):119–134. <https://doi.org/10.1007/s11747-015-0455-4>
- Ward P, McCreery J, Ritzman L, Sharma D (1998) Competitive priorities in operations management. *Decis Sci* 29(4):1035–1046. <https://doi.org/10.4236/jssm.2013.61008>
- Wilson P (1995) Detecting influential observations in data envelopment analysis. *J Product Anal* 6(1):27–45. <https://doi.org/10.1007/BF01073493>
- Winter S (2003) Understanding dynamic capabilities. *Strateg Manag J* 24(10):991–995. <https://doi.org/10.1002/smj.318>
- WorldBank (2019) New country classifications by income level: 2017–2018. <https://blogs.worldbank.org/opendata/new-country->

- [classifications-income-level-2017-2018](#). Accessed January 9, 2019
- Yang J (2008) Antecedents and consequences of knowledge management strategy: the case of Chinese high technology firms. *Prod Plan Control* 19(1):67–77. <https://doi.org/10.1080/09537280701776198>
- Zangwill W, Kantor P (1998) Toward a theory of continuous improvement and the learning curve. *Manage Sci* 44(7):910–920. <https://doi.org/10.1287/mnsc.44.7.910>
- Zhou H, Leong G, Jonsson P, Sum C (2009) A comparative study of advanced manufacturing technology and manufacturing infrastructure investments in Singapore and Sweden. *Int J Prod Econ* 120(1):42–53. <https://doi.org/10.1016/j.ijpe.2008.07.013>