International Journal of Production Research, 2018
Vol. 56, Nos. 1–2, 120–139, https://doi.org/10.1080/00207543.2017.1395489

Survival of the fittest: the impact of fit between warehouse management structure and warehouse context on warehouse performance

Nynke Faber[a], René B.M. De Koster[b]* and Ale Smidts[b]

[a]*Faculteit Militaire Wetenschappen, Netherlands Defence Academy, Breda, The Netherlands;* [b]*Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands*

(*Received 30 June 2017; accepted 6 October 2017*)

Warehouse operations are vital for the success of a supply chain. This paper explores fit among warehouse management structure and the context in which the warehouse operates as an important driver of warehouse performance. Warehouse management structure has been operationalised as the extensiveness by which warehouse operations are planned and controlled, and the complexity of the decision rules used for optimisation of the operations. Warehouse performance is measured using data envelopment analysis (DEA). Hypotheses are developed and tested in a survey study among 111 distribution warehouses in the Netherlands and Belgium. Our results show that indeed warehouse management structure should be adapted to context. In order to obtain high warehouse performance more complex warehouse tasks require more and more complex decision rules, and warehouses with more unpredictable demand require fewer tactical plans. These results can help warehouse managers to structure their warehouse management contingent on the context in which the warehouse operates.

Keywords: warehouse management; warehouse management structure; planning and control; performance; data envelopment analysis; cross efficiency

1. Introduction

Warehouses are important nodes in supply chains (Baker and Halim 2007). They play a critical intermediate role among supply chain members, affecting both supply chain costs and service (Kiefer and Novack 1999), particularly because distribution warehouses are often the final point in the supply chain for order assembly, value added services and dispatch to the customer (Baker and Halim 2007). Consequently, warehouse performance is pivotal for a supply chain's success and performance (Reiner and Hofmann 2006). The *International Journal of Production Research* is one of the key journals paying attention to research in warehouse design, management and performance (De Koster, Johnson, and Roy 2017). Warehouse performance measurement and comparison has been recognised as important by several authors (Kiefer and Novack 1999; Hackman et al. 2001; De Koster and Balk 2008; Tompkins et al. 2010; Johnson and McGinnis 2011; Hedler Staudt et al. 2015). Also, research has been conducted on the effect of different warehouse characteristics, such as ownership, country of origin, region location, lay-out, size and level of automation, on warehouse performance (e.g. Hackman et al. 2001; De Koster and Balk 2008; Banaszewska et al. 2012; Andrejić, Bojović, and Kilibarda 2013). Likewise, Johnson and McGinnis (2011) examined the relationship between different warehouse operational practices (e.g. use of pick-to-light, use of barcoding, use of temporary labour) and warehouse performance. In this paper, instead of examining the effect of a number of independent warehouse characteristics or operational practices on warehouse performance, we study the performance effects of warehouse operations management, defined here as the coherent whole of formal decisions controlling operations, and denoted here by warehouse management (WM). Note that, in practice, warehouse operations management encompasses more than just formal decisions. For example, leadership and informal decision-making are constituting elements. However, we focus on unambiguous formal WM decisions as supported by a Warehouse Management Information System (WMS). Generally, WM concerns the planning, optimisation and control of warehouse operations (Ten Hompel and Schmidt 2007). The operations that need to be planned, optimised and controlled include inbound flow handling, product-to-location assignment, product storage, order-to-stock location allocation, order batching and release, order picking, packing, value-added logistics activities and shipment. The way in which WM is structured manifests itself in the way decisions are made about the material flow and the use

*Corresponding author. Email: rkoster@rsm.nl

of resources (space, equipment and labour) in a warehouse in an everyday context. We define WM *structure* here as the blueprint specifying the way in which WM processes, i.e. planning, optimising and controlling, are formally organised.

In this paper, we study how WM structure affects warehouse performance by developing and testing a model for structuring high-performance WM. In this, we follow Sousa and Voss (2008) by taking a contingency perspective on operations management, meaning the structure-performance relationship is context dependent. Common to all contingency approaches is the proposition that performance is a consequence of the fit between structure and context (Donaldson 2001). A contingency perspective employs a reductionist approach (Sinha and Van de Ven 2005), meaning that we have to decompose WM structure into its constituent elements. We use, therefore, the dimensions of WM structure identified by Faber, de Koster, and Smidts (2013), i.e. planning, decision rules used to optimise warehouse operations, and control, to operationalise WM structure.

We hypothesise that warehouse performance is positively affected by a fit between WM structure and its context. We conduct an empirical study using the survey method to examine the proposed model and its associated hypotheses. Although Faber, de Koster, and Smidts (2013) researched the context drivers of WM structure, they did not examine whether the fit between these drivers and WM structure indeed improves warehouse performance. The current paper explicitly investigates such fit, and tests its impact on performance, thereby extending the findings of Faber, de Koster, and Smidts (2013) and translating them to insights for managers. From a managerial perspective, developing an understanding of the relationships between warehouse performance, WM structure and its context helps firms in deciding on their own optimal model for planning and controlling warehouse operations, and offers them more and better insights into the requirements of the supporting WMS. We also contribute to warehouse research, by empirically showing that fit, as measured by the interaction between WM structure and warehouse context, leads to higher warehouse performance. Additionally, this study applies a contingency approach in contrast to the dominant universal two-dimensional view of warehouse aspect–performance relationships, assuming the warehouse aspect leads to superior performance for all types of warehouses (e.g. Johnson and McGinnis 2011).

The paper is structured as follows. In the next section, we provide the theoretical and conceptual background in support of our hypotheses, followed by sections describing how data from actual warehouse settings were collected, discussing the measures of the constructs of the study, and reporting the results of the data analysis, respectively. In Section 6, we discuss the results. The final section ends with conclusions and suggestions for future research.

2. Background and hypotheses

2.1 *Contingency theory*

The major theoretical view on organisational contingencies is contingency theory (Lawrence and Lorsch 1967; Thompson 1967). In its most rudimentary form, this theory states that organisations adapt their structures to maintain fit with changing contextual factors. Failure to attain a proper fit between structure and context results in inferior outcomes. Central to contingency theory is the concept of fit between structural and contextual characteristics of organisations (Donaldson 2001). Applying the contingency perspective in our study, we thus propose that the performance of a warehouse or distribution centre (DC) is dependent upon the fit between WM structure and warehouse context. Sousa and Voss (2008) state that contingency studies involve three types of variables: (1) contingency variables, which represent the context; (2) response variables, which represent the organisational or managerial actions taken in response to contingency factors; and (3) performance variables, which measure the effectiveness of the organisation. Performance variables in contingency studies are the dependent measures and represent specific aspects of effectiveness that are appropriate to evaluate the fit between structure and context variables for the situation under consideration. In the contingency literature, there is no undisputed way to measure structure and context. Blackburn (1982) states that given the number of proposed structural dimensions and the variety of their definitions, identifying a definitive set of organisational dimensions or managerial actions is difficult without its specific context and objectives. This implies that each application of contingency theory should thus specify the structures that fit its contingency, so that fits and misfits are unique to that application (Donaldson 2001).

2.2 *Decomposing WM structure*

Faber, de Koster, and Smidts (2013) researched the drivers of WM structure. We specify the structural dimensions and contingency factors following that study. Formal operational decisions coordinate the material flows in and around the warehouse and the utilisation of the warehouse resources (space, equipment and labour) to satisfy customer demand. These decisions are the outcomes of the planning and control, and shop floor optimisation processes. Faber, de Koster, and Smidts (2013) decomposed WM structure into three structural dimensions: planning extensiveness, decision rules complexity and control sophistication. Planning extensiveness is related to the time and resources put into preparing tactical plans, such as stock, storage location assignment, transport and capacity (personnel and equipment) plans. Warehouses draw up tactical plans to make efficient use of resources and to fulfil market demand. Tactical plans define a framework for the daily shop floor planning level. Decision Rules Complexity refers to the complexity of shop floor decisions typically dealing with batching, sequencing, scheduling and routing of warehouse operations. Control is the process of coping with changes to plans and schedules, and control sophistication relates to the speed of the feedback and corrective action function of the management system.

2.3 *Warehouse contextual factors*

Faber, de Koster, and Smidts (2013) also concluded in their study that WM structure for DCs is contingent on two main warehouse contextual factors: task complexity and demand unpredictability. In the organisation science literature, there is consensus among researchers on two important organisational contextual factors: complexity (other terms used: variety, detail complexity or static complexity) and uncertainty (other terms used: environmental dynamism or dynamic complexity) (e.g. Duncan 1972; Premkumar and Zailani 2005; Van Assen 2005; Bozarth et al. 2009). Complexity is a consequence of the 'inner' boundary of the environment, i.e. the organisation itself, whereas uncertainty is a consequence of the external environment of the organisation. Complexity and uncertainty are an obvious first choice to include as main contingency factors in our study. We do acknowledge, however, that other factors may additionally play a role, e.g. management style, company culture, employee motivation or layout of the DC. In the current exploratory study, we operationalise complexity and uncertainty by task complexity and demand unpredictability, respectively. Task complexity is the number and diversity of tasks a DC has to perform and affects WM structure through the comprehensibility of the work to be done. Demand unpredictability refers to a warehouse's immediate environment that is uncontrollable by management and affects WM structure through the predictability of the work to be done. Faber, de Koster, and Smidts (2013) found the following context–structure relationships:

- The higher the task complexity, the more tactical plans are prepared (i.e. higher planning extensiveness).
- The higher the task complexity, the more, and the more complex decision rules are used to schedule and optimise warehouse operations (i.e. higher decision rules complexity).
- The higher the task complexity, the more sophisticated the control system is (i.e. higher control sophistication).
- The higher the demand unpredictability, the fewer tactical plans are prepared (i.e. lower planning extensiveness).

2.4 *Hypotheses and conceptual model of this study*

Contingency theory suggests that the effect of WM structure (i.e. levels of planning extensiveness, decision rules complexity and control sophistication) on warehouse performance depends on its contingencies (i.e. task complexity and demand unpredictability). Thus, fit of WM structure to contingencies leads to higher performance. This implies that when contingencies change, the WM structure should also change to fit the new level of the contingencies to avoid loss of performance. Therefore, we specifically hypothesise:

H1. Fit between Task Complexity and Planning Extensiveness will positively influence Warehouse Performance. Here fit means that a higher Task Complexity requires a higher Planning Extensiveness.

H2. Fit between Task Complexity and Decision Rules Complexity will positively influence Warehouse Performance. Here fit means that a higher Task Complexity requires a higher Decision Rules Complexity.

H3. Fit between Task Complexity and Control Sophistication will positively influence Warehouse Performance. Here fit means that a higher Task Complexity requires a higher Control Sophistication.

H4: Fit between Demand Unpredictability and Planning Extensiveness will positively influence Warehouse Performance. Here fit means that a higher Demand Unpredictability requires a lower Planning Extensiveness. Thus, Demand Unpredictability and Planning Extensiveness are negatively related to influence Warehouse Performance positively.

Also other warehouse factors may affect warehouse performance (in line with other studies, e.g. Banaszewska et al. 2012), independent of its context and structure. We distinguish three main control variables: industry sector, ownership (whether the DC is run by the company itself: insourced, or by a Logistic Service Provider: outsourced), and operations type (finished goods production DC, spare parts DC, wholesale DC, retail DC). Our full conceptual framework is shown in Figure 1.

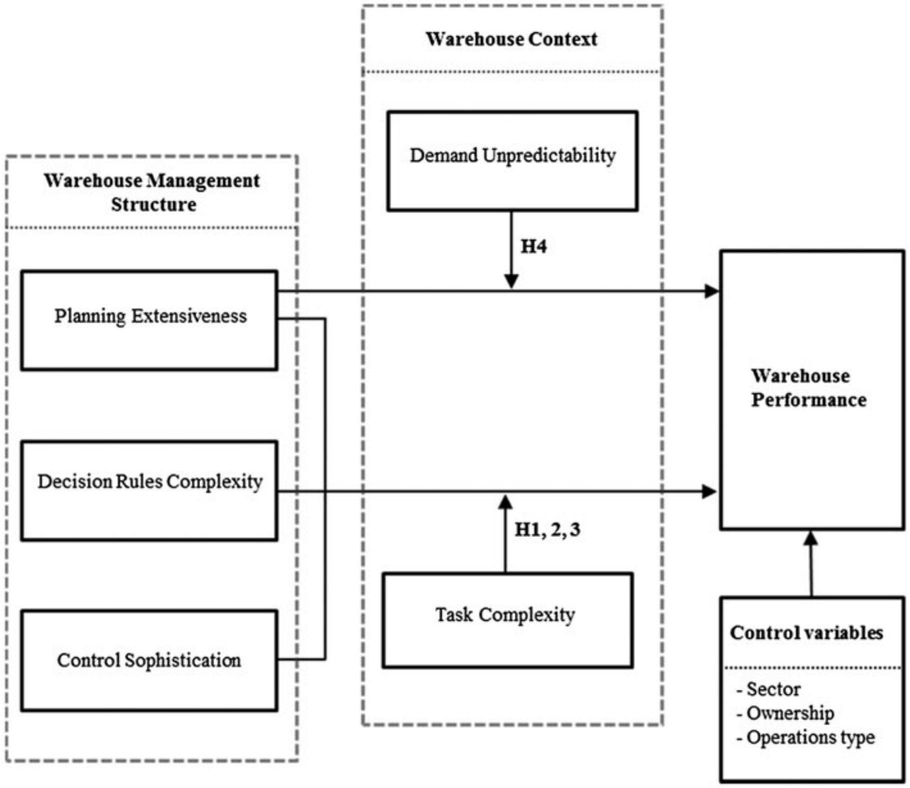

Figure 1. Conceptual framework of the performance implications of fit among WM structure and warehouse context.

3. Data collection and sample

To test our hypotheses, we use a subset of the database of 215 warehouses from Faber, de Koster, and Smidts (2013). They approached 765 warehouses in the Netherlands and Belgium by telephone and mail to fill out a questionnaire regarding warehouse context, WM structure and warehouse performance. The questions, response categories, scales and computation of the constructs are given in Appendices 1 and 2. Because the respondents represent an organisation and must be knowledgeable about the main constructs (Huber and Power 1985), a logistics or warehousing executive, preferably the warehouse manager, was requested to complete the questionnaire. In this study, we focus on warehouse performance data that were not reported in the research of Faber, de Koster, and Smidts (2013). We removed from the database 9 very small DCs (fewer than five direct FTEs – full-time equivalents and fewer than 1000 stock keeping units, or fewer than 60 order lines per day), and 21 production warehouses that supply raw materials to nearby production plants and do not focus on distribution. We also excluded DCs that could not (i.e. military and police DCs; $n = 8$) or did not answer all performance questions ($n = 66$). In this study, we, therefore, use data from 111 completed questionnaires. We tested non-response bias by comparing the subset of 111 DCs with the 66 DCs that did not respond to the performance questions across all variables of interest. A Mann-Witney test showed no significant differences at $p = 0.10$ level, except for decision rules complexity (DRC). On average, DRC is higher among subset 111. Although this research is limited to the Netherlands and Belgium, distribution practices in both countries are not different from elsewhere in Western Europe. In fact, many companies run multiple similar facilities in Western Europe (Quak and De Koster 2007, 1104). Since the database used by Faber, de Koster, and Smidts (2013) is sufficiently representative and the subset of 111 DCs does not differ from that database with respect to WM structure or warehouse context, we conclude that the current response is appropriate to draw meaningful conclusions. The average number of full-time (FTE) direct employees in the sample is 72 (SD = 85), the average number of stored SKUs per DC is 13,631 (SD = 23,396), and the average number of shipped order lines per day is 12,044 (SD = 23,886). Eighty-eight per cent of the respondents are senior warehouse managers and 12% are logistics staff members. See Table 1 for more descriptives of the sample.

Table 1. Sample descriptives (n = 111).

Ownership	%	Operations type	%	Sector	%	No. Direct FTEs	%	Size in m^2	%
DC insourced	63	Finished goods production DC	52.3	Automotive	6.3	< 10	10.8	< 1000	2.7
DC outsourced	37	Spare parts DC	6.3	Healthcare and Pharmaceutical	4.5	11–20	15.3	1000–3000	4.5
		Wholesale DC	19.8	Food retail	4.5	21–30	14.4	3000–5000	9.9
		Retail DC	21.6	Agricultural/Food products	7.2	31–50	15.3	5000–10,000	18.9
				ICT	4.5	51–70	11.7	10,000–20,000	27.0
				Industrial products	19.8	71–100	8.1	20,000–50,000	27.0
				Other products (mainly consumer products)	35.1	101–160	13.5	>50,000	9.9
				Public warehouses (multiple categories)	18.0	161–250	8.1		
						251–600	2.7		
Total	100	Total	100	Total	100	Total	100	Total	100

4. Construct measures

4.1 *Measuring warehouse context and WM structure*

Measurements of the constructs task complexity (TC), demand unpredictability (DU), planning extensiveness (PE), decision rules complexity (DRC) and control sophistication (CS), are taken from Faber, de Koster, and Smidts (2013). A summary of the warehouse context and the WM structure constructs, and their measures can be found in Appendix 1.

4.2 *Measuring fit between WM structure and warehouse context*

The key concept in this study is fit between WM structure and warehouse context. We define this fit as the appropriateness of the level of planning extensiveness, decision rules complexity and control sophistication to the level of task complexity and demand unpredictability as found by Faber, de Koster, and Smidts (2013). In the current study, we hypothesise that fit between WM structure and warehouse context predicts warehouse performance in such a way that a higher fit leads to a higher performance. Venkatraman (1989) provides an overview of various forms of fit, statistical methods used for analysis and the implicit assumptions made in the theoretical formulation and empirical analysis. As recommended by Venkatraman (1989), we adopt the 'fit as moderation' perspective to measure fit, because of the high degree of specificity of the theoretical relationships and the criterion-specificity (i.e. warehouse performance) of the hypotheses of this study. The moderation perspective assumes that the impact a predictor variable (in this research: WM structure) has on an outcome variable (in this research: warehouse performance) is dependent on the level of a third variable, the moderator (in this research: warehouse context). A linear model is assumed such that the moderator determines the sign and magnitude of the linear effect of the predictor on the outcome. In 'fit as moderation' method, fit is measured by the cross product (i.e. interaction effect) of two variables. A statistically significant interaction term indicates that the two variables (in this research: WM structure and warehouse context) exhibit a fit, and that this fit influences a dependent variable (in this research: Warehouse Performance). In general, the following equation is tested:

$$Y = B_0 + B_1X + B_2Z + B_3XZ + \varepsilon,$$

where Y is the outcome variable, X the moderator, Z the predictor, XZ the interaction effect of X and Z, B_i are the unstandardised regression coefficients, and ε is the error term. The moderation hypothesis is supported if B_3 differs significantly from zero. A positive interaction term implies that an increase (decrease) in a warehouse context variable makes the slope of the structural variables in predicting warehouse performance more positive (negative). In line with hypotheses 1, 2, 3 and 4, respectively, we test the impact of fit on performance by four interaction terms, each representing the cross-product of the standardised scores of:

(1) task complexity and planning extensiveness (TC × PE).
(2) task complexity and decision rules complexity (TC × DRC).
(3) task complexity and control sophistication (TC × CS).
(4) demand unpredictability and planning extensiveness (DU × PE).

4.3 *Measuring warehouse performance*

In general, warehouses aim at simultaneously reducing cost, increasing productivity and improving customer responsiveness (see the review paper of Hedler Staudt et al. 2015). Measuring warehouse performance provides feedback about how the warehouse performs compared to the requirements, or compared to peers. As such, it can also provide feedback on the adequacy and effectiveness of an implemented WM structure. Johnson and McGinnis (2011) discuss two types of warehouse operations performance criteria: financial (i.e. revenue related to cost), and technical (i.e. outputs related to inputs). They argue that technical criteria – based on output generated and resources consumed – tend to give a clearer picture of a warehouse's operational performance than financial measures, because warehouses typically do not generate revenues. Also Hedler Staudt et al. (2015) found in their literature review on warehouse performance measurement fewer works using cost related performance indicators than other operational performance indicators (time, quality and productivity). As warehouses are often part of a larger supply chain, traditional operational performance objectives including productivity, quality, delivery and flexibility (Schmenner and Swink 1998; Boyer and Lewis 2002) are more applicable. Technical performance measurement in the warehouse industry include cases or order lines picked per person per hour, picking or shipment errors rates, order throughput times and percentage of orders with special requests (Forger 1998; Van Goor, Ploos van Amstel, and Ploos van Amstel 2003). The problem with these indicators is that they are not mutually independent and that each of them depends on multiple input indicators (De Koster and Balk 2008). For example, the number of order lines picked per person per hour may be strongly influenced by system automation, assortment size and warehouse size. To overcome this problem, in this study, Data Envelopment Analysis (DEA) (Charnes, Cooper, and Rhodes 1978) is employed. DEA is a non-parametric linear programming technique that is capable of capturing all the relevant inputs (resources) and outputs into a single score of performance. It is probably the most commonly used method to measure warehouse performance in literature, used by e.g. Andrejić, Bojović, and Kilibarda (2013), Banaszewska et al. (2012), De Koster and Balk (2008), Hackman et al. (2001), and Johnson and McGinnis (2011). DEA measures the relative efficiency (performance) of a set of comparable decision-making units (DMU: the organisations under examination, e.g. warehouses). Fried, Lovell, and Schmidt (2008) provide a partial list of the many applications of DEA.

When determining the necessary input and output factors, all important aspects that determine the operational efficiency must be included. In the literature, different input–output models have been developed to benchmark warehouse operations. Most authors agree that the core inputs are labour, size (cost of space) and equipment (technology), representing the resources. With respect to the outputs, consensus only seems to exist on produced order lines.

In this study, we base our selection of input and output factors on De Koster and Balk (2008). The input factors of their model are labour, size, and equipment and the output factors are production output, quality and flexibility. They operationalise equipment by the degree of automation and the number of different stock keeping units (SKUs). The number of inputs and outputs is related to the number of observations. Simar and Wilson (2008) state that as the number of outputs increases the number of observations must increase at an exponential rate. Because of the relatively small size of the sample used in this study, we include only two indicators for production output: produced order lines and number of special operations where value added logistics (used by De Koster and Balk (2008) as separate output) is interpreted as a special operation. Furthermore, we combine quality and order lines in one output factor: effective (i.e. faultless) order lines. The input–output model we use in our study is shown in Figure 2. A compilation of the DEA input and output variables and their measures are summarised in Appendix 2.

In a DEA, positivity and isotonicity conditions must be met, which means that an increase in an input should increase one or more outputs (Bowlin 1998). Table 2 presents the Pearson correlation coefficients between the input and output factors of our DEA model. Although some of the input and output variables are measured at an ordinal scale, the number of classes is quite large (6 to 9; see Appendix 2 and Table 2) and we have therefore interpreted them on an interval scale in the computation. In spite of the fact that SKUs are significantly correlated to only one output factor (i.e. Flexibility, $p < 0.10$), all relationships of SKUs to outputs have the correct sign and we therefore maintain the model.

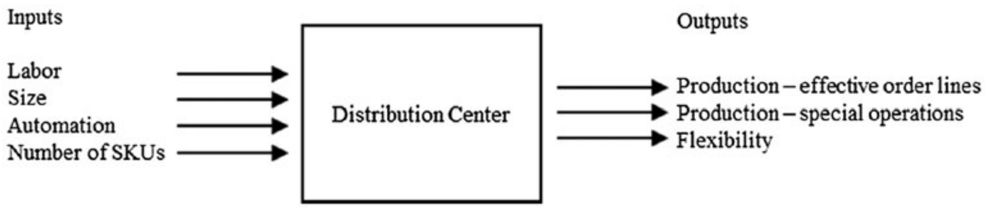

Figure 2. Input-output model of a distribution centre.

Table 2. Means, SDs, and Pearson correlation coefficients between DEA input and output variables (n = 111).

	Mean	SD	Effective order lines	Special operations	Flexibility
Labour (# FTEs)	72.4	85.3	0.74***	0.20**	0.14
Size (1–7 categories)	4.8	1.4	0.29***	0.26***	0.25***
Automation (1–6 categories)	3.7	1.3	0.33***	0.18*	0.04
SKUs (#)	13,631.4	23,395.8	0.14	0.12	0.16*
Effective Order lines (#)	11,830.7	23,305.0			
Special operations (#)	5.4	2.3			
Flexibility (3–9 categories)	7.4	1.1			

Note: Significant at: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

We use the original Charnes, Cooper, and Rhodes (1978) input-oriented constant-return-to-scale (CRS) approach for the DEA calculations (see Appendix 3). The maximum efficiency score is 100%, which means that the DC is efficient. An inefficient DC has an efficiency score between 0 and 100%. Input orientation of the model means that an inefficient DC with a score of x% should be able to achieve its output with only x% of its input resources. Constant-return-to-scale means that an increase in input results in a proportionate increase in output. However, we expect scale effects to play a role in warehouses, particularly in the size of the warehouse, as larger warehouses will not lead to proportionally more output. Therefore, in order to accommodate for scale effects in the DEA analysis, we measured the DEA-input variable 'Size' by seven non-linear categories (see Appendix 2). The DEA results indicate that 16% of the DCs (i.e. 18 DCs) operate efficiently. The mean efficiency score for the sample of this study is 0.65 with a standard deviation of 0.21. For completeness, we also executed an input-oriented variable-return-to-scale (VRS) model. VRS allows for scale effects. The difference between both models (CRS and VRS) appears to be limited: the mean ratio of CRS to VRS rates (i.e. scale efficiency) is 0.89 and SD is 0.11.

Unfortunately, the basic DEA model has some limitations. First, it does not distinguish between efficient DCs. Consequently, the distribution of efficiency scores is highly skewed. Second, it allows for unrestricted weight flexibility, which may result in identifying a DC with an unrealistic weighting scheme to be efficient (Eren Akyol and De Koster 2013). Such DCs perform well with respect to few input/output measures, but do not or hardly act as peer to other DCs.

In order to overcome these limitations, Sexton, Silkman, and Hogan (1986) propose the Cross Efficiency (CE) Evaluation model, which can identify good overall performers and distinguish between efficient DCs. In addition, because a DMU's score depends on all other peers, 'maverick' DMUs (those with very low weights for some key inputs and outputs) are penalised and receive a relatively lower score (Doyle and Green 1994). The CE model calculates the efficiency of each DC using the optimal input and output weights of all other DCs obtained from the DEA model. A 111 × 111 Cross Efficiency Matrix (CEM) is constructed using the cross efficiencies of all DCs. In the CEM, the element in the ith row and the jth column represents the efficiency of DC_j evaluated with respect to the optimal weights of DC_i. The elements in the diagonal of the CEM consist of DEA efficiencies, whereas the remaining elements represent the cross efficiency values. A DC with high efficiency values along its column is a good overall performer; a DC with low efficiencies along its column is a poor performer. Warehouse Performance (WP) in this study is measured by the average value of each column of the CEM. In this respect, CE is a robust DEA measure. See Appendix 3 for the so-called *aggressive* CE model (Sexton, Silkman, and Hogan 1986; Doyle and Green 1994; Liang et al. 2008) we use in this research and which gives unique weights and thereby reproducible results. According to Balk et al. (2017), the Sexton et al. implementation of cross-efficiency scores best compared to several other (cross-)efficiency measures on criteria like (1) methodological proximity to DEA, (2) ease of implementation, (3) extendibility, (4) discriminatory properties, (5) sensitivity to scale changes (6) sensitivity to erroneous data and (7) sensitivity to dominant DMU elimination. The average WP score for the sample is 0.40 with a standard deviation of 0.14. The minimum score is 0.17 and the maximum score is 0.85.

To check robustness of the WP measure, we also implemented super-efficiency scores (Lovell and Rouse 2003; Zhu 2001). The difference between the super-efficiency scores and the CE scores turned out to be minimal (Pearson correlation coefficient is 0.90).

5. Analysis and findings

5.1 *Regression analysis*

The objective of our study is to assess the effect of fit between WM structure and warehouse context on WP. Table 3 shows the correlations of the warehouse context variables, WM structure variables, the hypothesised interaction terms and WP.

To test our four hypotheses, we use a linear regression model. Note that using linear regression on DEA performance might have statistical issues (Simar and Wilson 2007, 2011). Not only because of the truncated nature of the dependent variable (requiring, e.g. Tobit regression), but also because of serial correlations between DEA scores (a slight modification of the score of a DMU on the efficient frontier may change the scores of other DMUs). Fortunately, CE scores suffer less from these issues (see Balk et al. 2017). WP is first regressed on the control variables (industry sector, ownership and operations type) stepwise. Next, the warehouse context variables (task complexity (TC) and demand unpredictability (DU)), WM structure variables (planning extensiveness (PE), decision rules complexity (DRC) and control sophistication (CS)) and hypothesised interaction terms (TC × PE, TC × DRC, TC × CS and DU × PE) are entered into the regression model. Table 4 shows the results. All reported *p*-values are two-tailed. Standardised explanatory (mean = 0 and standard deviation = 1) variables are employed in the regression model to ensure that differences in scale among the variables do not affect the results, and to increase interpretability of the regression terms. To check robustness of the regression results, we also use bootstrapping as recommended by Simar and Wilson (2007, 2011) to mitigate serial correlations. Bootstrapping appears to hardly impact the *p*-values (see last column in Table 4), except for the sector 'agriculture/food products' dummy variable. The overall pattern of the OLS regression results appears to be robust and is statistically valid to draw meaningful conclusions.

With respect to the control variables, the regression results in Table 4 show only significant effects for the industry sector 'agriculture/food products' on performance. Upon a more close inspection, the 'agricultural/food products' DCs

Table 3. Min, max, means, SDs and Pearson correlation coefficients of WM structure variables, warehouse context variables, interaction terms and warehouse performance.

	TC	DU	PE	DRC	CS	TC × PE	TC × DRC	TC × CS	DU × PE	WP
Task complexity (TC)	1									
Demand unpredictability (DU)	−0.18	1								
Planning extensiveness (PE)	0.33***	−0.27**	1							
Decision rules complexity (DRC)	0.49***	−0,19*	0.32***	1						
Control sophistication (CS)	0.23*	0.01	0.11	0.36***	1					
Interaction TC × PE	0.09	0.03	−0.26**	−0.04	0.03	1				
Interaction TC × DRC	0.05	0.00	−0.04	0.15	0.11	0.29**	1			
Interaction TC × CS	0.02	−0.02	0.03	0.10	0.10	0.04	0.27**	1		
Interaction DU × PE	0.03	−0.02	−0.01	0.07	0.05	−0.12	0.02	0.05	1	
Warehouse performance (WP)	−0.39***	−0.02	−0.15	−0.27**	−0.22*	−0.10	0.16	0.00	−0.14	1
n	111	110	111	111	110	111	111	110	110	111
Min	−2.26	−1.63	−2.32	−1.88	−1.67	−2.9	−2.7	−2.7	−3.9	0.16
Max	2.0	1.8	1.1	2.32	1.6	2.7	3.4	3.4	2.5	0.85
Mean	0	0	0	0	0	0	0	0	0	0.39
SD	1	1	1	1	1	1	1	1	1	0.14

Note: Significant at: **p* < 0.05; ***p* < 0.01; ****p* < 0.001.

Table 4. Regression results for warehouse performance.

	Warehouse Performance								
	Step 1				Step 2				Step 2 (based on 1000 bootstrap samples)
	B (unstandardised coefficient)	β (standardised coefficient)	t	p	B (unstandardised coefficient)	β (stanardised coefficient)	t	p	p
Constant	0.382	–	28.85	0.00**	0.375	–	25.47	0.00**	0.00**
Control variables									
Sector Agricultural/Food products	0.202	0.37	4.14	0.00**	0.125	0.23	2.27	0.03**	0.11
Warehouse characteristics									
Task Complexity (TC)					−0.030	−0.21	−1.95	0.05*	0.08*
Demand Unpredictability (DU)					−0.017	−0.12	−1.29	0.20	0.21
Warehouse Management variables									
Planning Extensiveness (PE)					−0.021	−0.15	−1.50	0.14	0.14
Decision Rules Complexity (DRC)					−0.012	−0.08	−0.76	0.45	0.47
Control Sophistication (CS)					−0.011	−0.08	−0.81	0.42	0.43
Interactions									
TC × PE					−0.021	−0.14	−1.41	0.16	0.17
TC × DRC					0.030	0.18	1.92	0.06*	0.08*
TC × CS					−0.005	−0.04	−0.39	0.69	0.68
DU × PE					−0.020	−0.15	−1.78	0.08*	0.10*
R^2	0.14				0.29				
F-value	17.109				2.370				
Sig F Change	0.00				0.02				
R^2 change	0.14				0.15				
n	109				109				

Notes: Significant at: *$p < 0.10$, **$p < 0.05$.

($n = 8$) appear to have a significant lower TC than the other sectors, which might explain the relatively high performance of this group. No differences in performance are found between insourced and outsourced DCs, and between different operation types.

In step 2, an omnibus *F*-test shows that the added variables of interest (i.e. main effects and interactions) representing our four hypotheses contribute significantly to the variance explained over and above the first step ($F = 2.37$, $p < 0.02$; R^2-change = 0.15). In step 2, a statistically significant interaction term would indicate that the two variables exhibit a fit, and that this fit influences the independent variable (WP). This applies to *H2* with interaction term TC × DRC ($\beta = 0.18$; $p = 0.06$) and *H4* with interaction term DU × PE ($\beta = -0.15$; $p = 0.08$) but not for *H1* with interaction term TC × PE ($\beta = -0.14$; $p = 0.16$) and *H3* with interaction term TC × CS ($\beta = -0.04$; $p = 0.69$). The interaction term DU × PE is negative when there is fit between DU and PE: a higher DU requires a lower PE. Therefore, in Table 4, the interaction term DU × PE has the expected sign by influencing WP negatively. Furthermore, Table 4 shows a significant negative main effect of TC on WP ($\beta = -0.21$; $p = 0.05$).

5.2 *Interpreting the effects of fit on WP*

The direction of the effect of fit cannot be interpreted solely from the β-coefficient of the interaction term because the main effects (single variable terms) and interaction term must be interpreted collectively (Venkatraman 1989; Hoffman et al. 1992; Stock and Tatikonda 2008). We, therefore, need to delve a bit deeper to find out the nature of the moderation (Aiken and West 1991; Stock and Tatikonda 2008; Hayes 2013; Dawson 2014). First, we discuss interaction effect TC × DRC, testing *H2*. It is common to plot the effect to improve its interpretability. A simple slope is defined as the regression of the outcome (WP) on the predictor (DRC) at a specific value of the moderator (TC). We select TC low and TC high values to be the mean minus one standard deviation, and the mean plus one standard deviation, respectively. Because the variables are standardised, the low and high TC values are −1 and + 1, respectively. Figure 3 shows the respective regression lines. WP is higher when there is a greater fit between TC and DRC. This can be seen best by examining the endpoints of the two lines in Figure 3.

For a high level of DRC (DRC = +2.32), there is a higher level of WP when TC is high (TC = +1) than when TC is low (TC = −1). For a low level of DRC (DRC = −1.88), there is a higher level of WP for the TC low line (TC = −1) than for the TC high line (TC = +1). To test whether the relationship between DRC and WP significantly changes at different levels of TC, we compare these slopes in a simple slopes analysis (Aiken and West 1991). Results show that when TC is low, there is a significant negative relationship between DRC and WP ($\beta = -0.03$; $p = 0.07$) (striped line in Figure 3), and when TC is high (solid line in Figure 3), there is a positive but non-significant relationship between DRC and WP ($\beta = +0.01$; $p = 0.42$). That is, when TC is low, implementing more decision rules and more complex ones significantly reduces WP. Implementing more, and more complex decision rules when TC increases has a positive effect on WP, but this effect is not significant. In conclusion, the results of the current study show that *H2* is supported when task complexity is low: with decreasing TC, a lower DRC is required to achieve a higher WP. However, *H2* is only directionally supported when Task Complexity is high: with increasing TC, a higher DRC tends to increase WP.

Second, we discuss the interaction effect DU × PE, testing *H4*. We again select DU low and DU high values to be the mean minus one standard deviation, and the mean plus one standard deviation, respectively. Because DU is

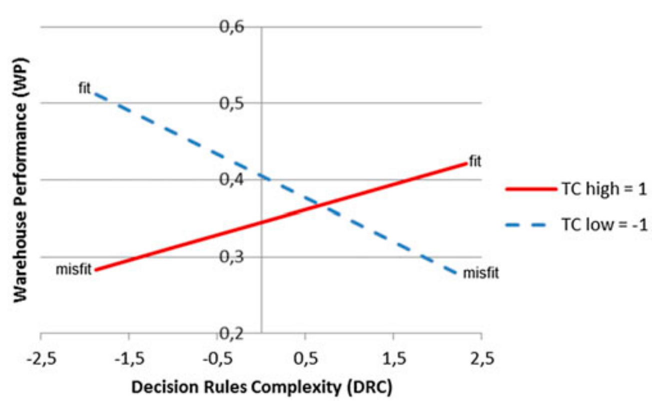

Figure 3. Interaction effect TC × DRC on warehouse performance.

standardised, low and high DU values are −1 and +1, respectively. Figure 4 shows that as PE increases, the WP score is higher when there is a better fit between DU and PE. This can be seen by examining the endpoints of the two regression lines shown in Figure 4.

For a high level of PE (PE = 1.1), WP is higher for DU low (DU = −1) than for DU high (DU = +1). For a low level of PE (PE = −2.32), WP is higher for DU high (DU = 1) than for DU low (DU = −1). Applying a simple slopes analysis reveals that when DU is low, there is no relationship between PE and WP ($\beta = -0.01$; $p = 0.74$) (striped line in Figure 4), and when DU is high (solid line in Figure 4), there is a significant negative relationship between PE and WP ($\beta = -0.04$; $p = 0.02$). In conclusion, the data does not support the relationship between PE and WP when DU is low but, for a high level of DU (i.e. demand is difficult to predict), more extensive planning influences WP significantly negative. Thus, the results show that *H4* is only supported for high DU: to achieve a high performance, a higher DU requires a less extensive planning. No support was found for *H1* and *H3*.

5.3 *Additional tests*

In order to test for robustness of our results, we also regressed on alternative measures of WP, both super-efficiency (Zhu 2001; Lovell and Rouse, 2003) and labour productivity (effective order lines per direct FTE). Although super-efficiency is sensitive to data outliers by nature (Zhu 2001), it gives about the same results for the interaction variables (TC × DRC; $\beta = 0.19$, $p < 0.06$, and DU × PE; $\beta = -0.18$, $p < 0.04$). Using labour productivity as measure of WP lowers the explanatory power of the model significantly, as it only looks at a limited aspect of overall performance, leading to non-significant interaction effects. In conclusion, CE DEA is a robust measure for performance and by its use of aggregate scores, it is not very sensitive to data errors.

6. Discussion of empirical results

First, we discuss the effect of planning structure on performance. The planning–performance relationship was hypothesised to be contingent on Task Complexity as well as on Demand Unpredictability, but with conflicting implications (see *H1* and *H4* respectively). The results exemplified in Figure 4 show that when demand is more difficult to predict, performance indeed increases if planning efforts are limited. In other words, more extensive tactical planning does not help to improve performance when demand is more unpredictable. Most likely this is because resources are captured in drawing up and maintaining plans without generating proportionally better quality output, because demand forecasts change constantly. De Koster and Shinohara (2006) found something similar in their multiple case study: companies putting a lot of effort into services do not necessarily perform better (in multifactor performance), as the extra effort may not pay off proportionally. When demand is predictable, we found no effect of planning on performance. It seems that performance is indifferent to putting more effort into planning when demand is predictable. Also, the results show no moderation effect of task complexity on the planning–performance relationship. We considered that not finding a significant moderation effect (TC × PE) might be affected by demand unpredictability. Probably, extensive planning makes sense when a warehouse task is complex *and* demand is predictable, because both effects add up. On the other hand, extensive planning does not improve performance when demand is unpredictable, regardless of whether a warehouse task is simple or complex. In addition, when a warehouse task is simple and demand is predictable, the

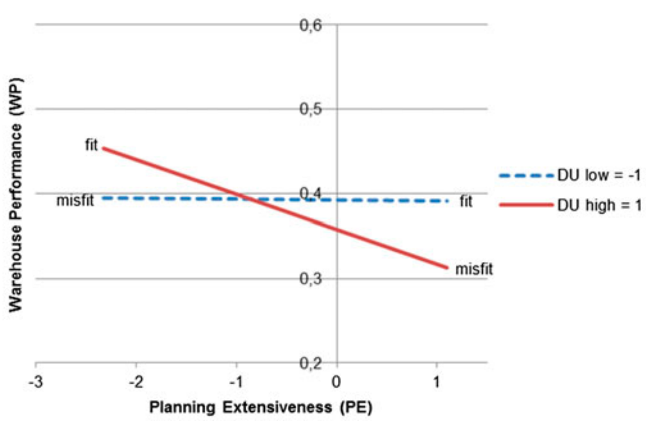

Figure 4. Interaction effect DU × PE on warehouse performance.

planning-performance relationship is unclear, but extensive planning seems out of proportion wasting time and money. However, a three-way interaction term (TC × DU × PE) added to the regression in Table 4 did not support these suggestions. This may be due to the relatively small sample size of 111 DCs, which greatly reduces the power of the test.

Second, we discuss the effect of decision rules structure on performance (*H2*). The research results (see Figure 3) show that managing a simple warehouse task by applying a limited number of simple decision rules increases performance. This implies that in this case, investing in software offering many different and complex decision rules seems to be a waste of money. Furthermore, the results show that managing a more complex warehouse task by applying more, and more complex decision rules affect performance positively, even though in our empirical study the effect is only directionally significant.

Third, we found no moderating effect of task complexity on the relationship between the control element of WM structure and warehouse performance (*H3*). This may be due to the current operationalisation of control sophistication that solely focused on swift (online or real time) data processing to allow human decision-making. Probably, human behaviour, especially with respect to interpreting information and taking action on it, is another important dimension of the control system, which was not considered in this research.

All in all, the results lead to the model summarised in Figure 5 indicating how to structure high performance WM for different situations. A simple WM structure (i.e. limited planning efforts, and limited number of simple decision rules) increases performance when the warehouse task is simple and demand is difficult to predict (see first row in Figure 5). A simple WM structure is also expected to increase performance when the task is simple and demand is predictable (see third row in Figure 5), but this is only partly supported by this research. Furthermore, a WM structure that typically focuses on optimisation (i.e. limited planning efforts, and more plus more complex decision rules) increases performance, although not significant, when the warehouse task is complex and demand is difficult to predict (see second row in Figure 5). Finally, a complex WM structure (i.e. extensive planning, and more plus more complex decision rules) is expected (but not supported by the research) to increase performance when the task is complex and demand is predictable (see last row in Figure 5).

Contingencies (warehouse context)	Value	Warehouse Management structure	Warehouse Performance
Task Complexity	low	Simple (i.e., limited planning efforts + limited number of simple decision rules)	High
Demand Unpredictabillity	high		
Task Complexity	high	Optimization focused (i.e., limited planning efforts + more, and more complex decision rules)	High *(tendency)*
Demand Unpredictabillity	high		
Task Complexity	low	Simple (i.e., limited planning efforts + limited number of simple decision rules)	High *(expected)* *Partly supported by this research*
Demand Unpredictabillity	low		
Task Complexity	high	Complex (i.e., extensive planning + more, and more complex decision rules)	High *(expected)* *Not supported by this research*
Demand Unpredictabillity	low		

Figure 5. Confirmed model for structuring high-performance WM.

7. Conclusions and future research

This paper takes a first step towards understanding the impact of WM structure on warehouse performance. The study shows that the appropriate levels of planning extensiveness and decision rules complexity indeed explain a significant portion of variance in performance (29%). This research contributes to the body of knowledge in the field of warehouse management by applying contingency theory and empirically assessing the impact of the degree of fit between WM structure and warehouse context on warehouse performance. Second, it contributes to the knowledge of structuring WM, as expressed in Figure 5.

Our results are also of importance for managers, as they show that demand unpredictability and task complexity can effectively be managed by choosing the appropriate level of planning and level of decision rules complexity. Managers can also use this knowledge in selecting appropriate planning and control systems (e.g. WMS) for their warehouse, fitting the context. Warehouse planning systems that are too extensive in unpredictable contexts or scheduling and optimisation that are too complex in simple contexts imply a misfit and lead to underperformance and waste of money.

Several avenues for further research exist. Follow up research could focus on refining the measures of WM structure and in particular to comprehend the joint effects of task complexity and demand unpredictability on planning. Also, a larger sample of warehouses would enable testing the proposed relationships with more statistical power. In the current study, we found directional support for some of our hypothesised effects which might be confirmed in a larger sample. As remarked by Simar and Wilson (2007) and Johnson and Kuosmanen (2011, 2012), using OLS regression with a DEA score as dependent variable has statistical limitations. Although we mitigate the effects using CE scores, a future study might include alternative regression models, such as truncated regression (Simar and Wilson 2007, 2011) or concave nonparametric least squares regression (Johnson and Kuosmanen 2011, 2012). Furthermore, the interpretation of performance measured using DEA and then CE is fairly complex (Anderson, Hollingsworth, and Inman 2002). Therefore, also other research methods than survey research would be helpful for validating the model. For example, the impact of fit between structure and context could be tested by means of action research (Coghlan and Brannick 2014). In such a study performance effects, e.g. measured by different ratios, could be studied as structure elements are adapted to the warehouse context. Also, the proposed model could be further validated by means of targeted case studies. For example, high and low performing warehouses could be selected and compared with respect to their degree of fit among structure and context. The explanation of why particular combinations of context and structure occur in distribution centres is solely based on Faber, de Koster, and Smidts (2013). Therefore, more research is called for to more fundamentally understand context and structure relationships in distribution centres, e.g. by semi-structured interviews with warehouse managers.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Aiken, L. S., and S. G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, California: Sage.

Anderson, T. R., K. Hollingsworth, and L. Inman. 2002. "The Fixed Weighting Nature of a Cross-Evaluation Model." *Journal of Productivity Analysis* 17 (3): 249–255.

Andrejić, M., N. Bojović, and M. Kilibarda. 2013. "Benchmarking Distribution Centres Using Principal Component Analysis and Data Envelopment Analysis: A Case Study of Serbia." *Expert Systems with Applications* 40 (10): 3926–3933.

Baker, P., and Z. Halim. 2007. "An Exploration of Warehouse Automation Implementations: Cost, Service and Flexibility Issues." *Supply Chain Management: An International Journal.* 12 (2): 129–138.

Baker, R. C., and S. Talluri. 1997. "A Closer Look at the Use of Data Evenlopment Analysis for Technology Selection." *Computers & Industrial Engineering* 32 (1): 101–108.

Balk, B., R. de Koster, C. Kaps, and J. Zofio. 2017. *What is Cross-efficiency? A Comparison of Performance Benchmarking Methods*. Working paper. Erasmus University Rotterdam.

Banaszewska, A., F. Cruijssen, W. Dullaert, and J. C. Gerdessen. 2012. "A Framework for Measuring Efficiency Levels – The Case of Express Depots." *International Journal of Production Economics.* 139 (2): 484–495.

Blackburn, R. S. 1982. "Dimensions of Structure: A Review and Reappraisal." *The Academy of Management Review.* 7 (1): 59–66.

Bowlin, W. F. 1998. "Measuring Performance: An Introduction to Data Envelopment Analysis (DEA)." *The Journal of Cost Analysis.* 15 (2): 3–27.

Boyer, K. K., and M. W. Lewis. 2002. "Competitive Priorities: Investigating the Need for Trade-Offs in Operations Strategy." *Production and Operations Management.* 11 (1): 9–20.

Bozarth, C. C., D. P. Warsing, B. B. Flynn, and E. J. Flynn. 2009. "The Impact of Supply Chain Complexity on Manufacturing Plant Performance." *Journal of Operations Management.* 27 (1): 78–93.

Charnes, A., W. W. Cooper, and E. Rhodes. 1978. "Measuring the Efficiency of Decision-making Units." *European Journal of Operational Research* 2 (6): 429–444.

Coghlan, D., and T. Brannick. 2014. *Doing Action Research in Your Own Organization*. London: Sage.

Dawson, J. F. 2014. "Moderation in Management Research: What, Why, When and How." *Journal of Business and Psychology* 29 (1): 1–19.

De Koster, M. B. M., and B. M. Balk. 2008. "Benchmarking and Monitoring International Warehouse Operations in Europe." *Production and Operations Management* 17 (2): 175–183.

De Koster, R., and M. Shinohara. 2006. "Supply-chain Culture Clashes in Europe. Pitfalls in Japanese Service Operations." *Supply Chain Forum: An International Journal* 7 (1): 60–68.

De Koster, R., A. L. Johnson, and D. Roy. 2017. "Warehouse Design and Management." *International Journal of Production Research* 55 (21): 6407–6422.

Donaldson, L. 2001. *The Contingency Theory of Organizations*. Thousand Oaks, California: Sage.

Doyle, J., and R. Green. 1994. "Efficiency and Cross-efficiency in DEA: Derivations, Meanings and Uses." *Journal of the Operations Research Society* 45 (5): 567–578.

Duncan, R. 1972. "Characteristics of Organizational Environments and Perceived Environmental Uncertainty." *Administrative Science Quarterly* 17 (3): 313–327.

Eren Akyol, D., and R. B. M. De Koster. 2013. "Non-dominated Time-window Policies in City Distribution." *Production and Operations Management* 22 (3): 739–751.

Faber, N., M. B. M. de Koster, and A. Smidts. 2013. "Organizing Warehouse Management." *International Journal of Operations and Production Management* 33 (9): 1230–1256.

Forger, G. 1998. "Benchmark Your Warehouse for Future Success." *Modern Materials Handling.* 53 (12): 39–41.

Fried, H. O., C. A. K. Lovell, and S. S. Schmidt. 2008. *The Measurement of Productive Efficiency and Productivity Change*. New York: Oxford University Press.

Hackman, S. T., E. H. Frazelle, P. M. Griffin, S. O. Griffin, and D. A. Vlasta. 2001. "Benchmarking Warehousing and Distribution Operations: An Input–Output Approach." *Journal of Productivity Analysis* 16 (1): 79–100.

Hayes, A. F. 2013. *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression Based Approach*. New York: Guilford Press.

Hedler Staudt, F., G. Alpan, M. Di Mascolo, and C. M. Taboada Rodriguez. 2015. "Warehouse Performance Measurement: A Literature Review." *International Journal of Production Research* 53 (18): 5524–5544.

Hoffman, J. J., J. B. Cullen, N. M. Carter, and C. F. Hofacker. 1992. "Alternative Methods for Measuring Organization Fit: Technology, Structure, and Performance." *Journal of Management* 18 (1): 45–57.

Huber, G. P., and D. J. Power. 1985. "Retrospective Reports of Strategic-level Managers: Guidelines for Increasing Their Accuracy." *Strategic Management Journal.* 6 (2): 171–180.

Johnson, A. L., and T. Kuosmanen. 2011. "One-stage Estimation of the Effects of Operational Conditions and Practices on Productive Performance: Asymptotically Normal and Efficient, Root-N Consistent StoNEZD Method." *Journal of Productivity Analysis* 36 (2): 219–230.

Johnson, A. L., and T. Kuosmanen. 2012. "One-Stage and Two-Stage DEA Estimation of the Effects of Contextual Variables." *European Journal of Operational Research* 220 (2): 559–570.

Johnson, A. L., and L. F. McGinnis. 2011. "Performance Measurement in the Warehousing Industry." *IIE Transactions* 43 (3): 220–230.

Kiefer, A. W., and R. A. Novack. 1999. "An Empirical Analysis of Warehouse Measurement Systems in the Context of Supply Chain Implementation." *Transportation Journal* 38 (3): 18–27.

Lawrence, P. R., and J. W. Lorsch. 1967. *Organization and Environment*. Boston, MA: Harvard Business Graduate School Press.

Liang, L., J. Wu, W. D. Cook, and J. Zhu. 2008. "Alternative Secondary Goals in DEA Cross-efficiency Evaluation." *International Journal of Production Economics* 113 (2): 1025–1030.

Lovell, C. A. K., and A. P. B. Rouse. 2003. "Equivalent Standard DEA Models to Provide Super-efficiency Scores." *Journal of Operational Research Society* 54 (1): 101–108.

Premkumar, R., and S. Zailani. 2005. "Supply Chain Integration and Performance: US versus East Asian Companies." *Supply Chain Management: An International Journal* 10 (5): 379–393.

Quak, H. J., and M. B. M. De Koster. 2007. "Exploring Retailers' Sensitivity to Local Sustainability Policies." *Journal of Operations Management* 25 (6): 1103–1122.

Reiner, G., and P. Hofmann. 2006. "Efficiency Analysis of Supply Chain Processes." *International Journal of Production Research* 44 (23): 5065–5087.

Schmenner, R. W., and M. L. Swink. 1998. "On Theory in Operations Management." *Journal of Operations Management.* 17 (1): 97–113.

Sexton, T. R., R. H. Silkman, and A. J. Hogan. 1986. "Data Envelopment Analysis: Critique and Extensions." In *Measuring Efficiency: An Assessment of Data Envelopment Analysis 32*, edited by R. H. Silkman, 73–105. San Francisco, CA: Jossey Bass.

Simar, L., and P. W. Wilson. 2007. “Estimation and Inference in Two-Stage, Semi-Parametric Models of Productive Efficiency.” *Journal of Econometrics* 136 (1): 31–64.

Simar, L., and P. W. Wilson. 2008. “Statistical Inference in Nonparametric Frontier Models: Recent Developments and Perspectives.” In *The Measurement of Productive Efficiency*. 2nd ed., Chap. 4, edited by H. Fried, C. A. K. Lovell, and S. S. Schmidt, 421–521. Oxford: Oxford University Press.

Simar, L., and P. W. Wilson. 2011. “Two-stage DEA: Caveat Emptor.” *Journal of Productivity Analysis* 36: 205–218.

Sinha, K. K., and A. H. Van de Ven. 2005. “Designing Work within and between Organizations.” *Organization Science* 16 (4): 389–408.

Sousa, R., and C. A. Voss. 2008. “Contingency Research in Operations Management Practices.” *Journal of Operations Management* 26 (6): 697–713.

Stock, G. N., and M. V. Tatikonda. 2008. “The Joint Influence of Technology Uncertainty and Interorganizational Interaction on External Technology Integration Success.” *Journal of Operations Management* 26 (1): 65–80.

Ten Hompel, M., and T. Schmidt. 2007. *Warehouse Management: Automation and Organisation of Warehouse and Order Picking Systems*. Berlin: Springer.

Thompson, J. D. 1967. *Organization in Action*. New York: McGraw-Hill.

Tompkins, J. A., J. A. White, Y. A. Bozer, and J. M. A. Tanchoco. 2010. *Facilities Planning*. New York: John Wiley & Sons.

Van Assen, M. F. 2005. “Empirical Studies in Discrete Parts Manufacturing Management.” PhD diss., RSM Erasmus University Rotterdam.

Van Goor, A. R., M. J. Ploos van Amstel, and W. Ploos van Amstel. 2003. *European Distribution and Supply Chain Logistics*. Groningen: Stenfert Kroese.

Venkatraman, N. 1989. “The Concept of Fit in Strategy Research: Toward Verbal and Statistical Correspondence.” *Academy of Management Review* 14 (3): 423–444.

Zhu, J. 2001. “Super-efficiency and DEA Sensitivity Analysis.” *European Journal of Operational Research* 129 (2): 443–455.

Appendix 1. Summary of measures of warehouse context and WM structure variables

Variable	Description	Accompanying question/ instruction	Response categories	Computation
TC	*Task complexity*			Sum (TCa, TCb, TCc)[a]
TCa	Log number of SKUs	What is the average number of SKU?	(open)	Log (number of SKUs)
TCb	Operations diversity			Sum (TCb1,TCb2)
TCb1	Number of special operations	Tick operation if applicable	• quality control • return handling • recoding • cross-docking • product repacking • cycle counting • internal product transportations • value added logistics • other special operations	Count number of operations ticked. 1: count is 5 or less 2: count is more than 5
TCb2	Average number of modes	What is the average number of modes in which operations can be carried out?	• 3 or less • more than 3	1: average number of modes per operation is 3 or less 2: average number of modes per operation is more than 3
TCc	Log number of order lines	What is the average number of order lines per day?	(open)	Log (number of order lines)
DU	*Demand unpredictability*			Sum (DU1, DU2, DU3)
DU1	Demand unpredictability long term	How predictable is the total number of order lines for long term (half a year – 1 year)?	• predictable • predictable to a limited extent • difficult to predict	1: demand is predictable 2: demand is predictable to a limited extent 3: demand is difficult to predict
DU2	Demand unpredictability short term	How predictable is the total number per product/product group for short term? (1 week – 1 month)	• predictable • predictable to a limited extent • difficult to predict	1: demand is predictable 2: demand is predictable to a limited extent 3: demand is difficult to predict
DU3	Demand unpredictability very short term (1 day)	How predictable is the total number of order lines per day for the very short term (1 day – 1 week)?	• predictable • predictable to a limited extent • difficult to predict	1: demand is predictable 2: demand is predictable to a limited extent 3: demand is difficult to predict
PE	*Planning Extensiveness*	Tick plans if applicable	• stock planning • storage location planning • capacity planning • transport planning	Count number of plans ticked
DRC	*Decision Rules Complexity*			Sum (Dca, DCb)[a]
DRCa	Number of activities explicitly using decision rules	Tick activities using decision rules	• Allocate dock doors to inbound transport units (e.g. trucks) • Allocate capacity (personnel and equipment) to inbound transport units • Allocate inbound products to storage locations or cross-docking	Count number of activities ticked

			• Allocate inbound products and storage locations to inbound routes (optimal routes) • Allocate inbound routes to available capacities • Cluster orders to be shipped to transport units • Cluster customer orders to batches • Settle the executing start point per batch • Allocate pick order lines to storage locations • Allocate pick order lines and storage locations to pick routes (optimal routes) • Allocate pick routes to available capacities • Allocate dock doors to outbound transport units	
DRCb	Perceived complexity of decision rules			Sum (DCb1, DCb2)
DRCb1	Perceived complexity of inbound decision rules	How complex are inbound decision rules?	• simple • not simple, not complex • complex	1: decision rules are simple 2: decision rules are not simple and not complex 3: decision rules are complex
DRCb2	Perceived complexity of outbound decision rules	How complex are outbound decision rules?	• simple • not simple, not complex • complex	1: decision rules are simple 2: decision rules are not simple and not complex 3: decision rules are complex
CS	*Control Sophistication*			Sum (CS1, CS2, CS3, CS4, CS5)
CS1	Operational data registration	Is operational data generally registered real time and online?	• no • yes	1: registration is not real time and online 2: registration is real time and online
CS2	Operational information availability	Is information for operational control generally available real time and online?	• no • yes	1: information is not real time and not available online 2: information is real time and available online
CS3	Reaction to unforeseen situations	What is the reaction to unforeseen situations?	• takes some time • immediate	1: reaction takes some time 2: reaction is immediate
CS4	Operational information from suppliers	Is operational information from suppliers (e.g. stock levels) available online?	• no • yes	1: supplier information is not available online 2: supplier information is available online
CS5	Operational information from customers	Is operational information from customers (e.g. point-of-sale information) received online?	• no • yes	1: customer information is not received online 2: customer information is received online

[a]All variables are standardised before summing.

Appendix 2. Summary of measures of DEA input and output variables

Variable	Description	Accompanying question/instruction	Response categories	Computation
Labour	*Number of direct FTEs*	What is the average number of FTEs working in the distribution centre?	(open) NB. The yearly number of working hours per full-time employee does not differ much in the Netherlands and Flanders due to national agreements with labour unions	Number of direct FTEs
Size	*Size of the distribution centre in m^2*	What is the size of the distribution centre?	• less than 1000 m^2 • 1000 to 3000 m^2 • 3000 to 5000 m^2 • 5000 to 10,000 m^2 • 10,000 to 20,000 m^2 • 20,000 to 50,000 m^2 • more than 50,000 m^2	1: size is less than 1000 m^2 2: size is between 1000 and 3000 m^2 3: size is between 3000 and 5000 m^2 4: size is between 5000 and 10,000 m^2 5: size is between 10,000 and 20,000 m^2 6: size is between 20,000 and 50,000 m^2 7: size is more than 50,000 m^2
Automation	*Level of automation*			Combination of 'Automated Systems' and 'Information System Specificity': 1: no automated systems are used 2: basic automation (ERP-system) 3: warehouse management system (WMS) 4: WMS, barcoding and wireless communication are used 5: WMS, barcoding, wireless communication and light technology systems (conveyor, pick-to-light, put-to-light) are used 6: WMS, barcoding, wireless communication and automated systems like automated-guided vehicles, miniloads, automatic storage and retrieval systems, sorters, or robots are used
Automation: 'Automated systems'	Different types of automated systems used in the distribution centre	Which automated systems are used in the distribution centre (e.g. cranes, sorters, palletisers, AGV, radio frequency)?	(open)	
Automation: 'Information System Specificity'	Information system used in the distribution centre	Tick information system type if applicable	• no automated information system • standard ERP • standard ERP with substantial customisation • standard WMS • standard WMS with substantial customisation • tailor-made system	0: no automated information system 1: standard ERP 2: standard ERP with substantial customisation 3: standard WMS 4: standard WMS with substantial customisation 5: tailor-made system
SKUs	*Number of SKUs*	What is the average number of stored SKUs?	(open)	Number of SKUs
Production – Effective order lines	*Number of effective order lines*			Multiplication of Production-order lines by Quality

Production – order lines	Number of order lines	What is the average number of order lines per day?	(open)	Number of order lines
Quality	The percentage of error-free order lines	What is the percentage of error-free order lines shipped? (errors include wrong number of products, wrong products, late or early delivery, wrong packaging, etc.)	(percentage)	Percentage of error-free order lines. 'No idea of this percentage'/ 'We do not measure' was given a low 90% score
Production – special operations	*Number of special operations*	Tick operation if applicable	• quality control • return handling • recoding • cross-docking • product repacking • cycle counting • internal product transportations-value added logistics • other special operations	Count number of operations ticked
Flexibility	*The perceptive ability to respond to changes in the market environment compared to competitors*			Sum (Flexibility 'Production Volume', Flexibility 'Changes', Flexibility 'Response')
Flexibility: 'Production Volume'	The ability to respond to changes in production volume	The ease to respond to changes in production volume compared to competitors	• worse than competitors • equal to competitors • better than competitors	1: worse than competitors 2: equal to competitors 3: better than competitors
Flexibility: 'Changes'	The ability to respond to late changes in orders	The ease to respond to late changes in orders compared to competitors	• worse than competitors • equal to competitors • better than competitors	1: worse than competitors 2: equal to competitors 3: better than competitors
Flexibility: 'Response'	The ability to respond quickly to orders	The ease to deliver orders faster compared to competitors	• worse than competitors • equal to competitors • better than competitors	1: worse than competitors 2: equal to competitors 3: better than competitors

Appendix 3. DEA and cross efficiency

The original DEA model (Charnes, Cooper, and Rhodes 1978) known as the CCR model, considers n *DMUs* each with m inputs and s outputs to be evaluated. The jth input and the kth output of DMU_o are denoted by x_{jo} where $j = 1, \ldots, m$ and y_{ko} where $k = 1, \ldots, s$, respectively. The ratio of the weighted combination of outputs to the weighted combination of inputs is used to measure the relative efficiency of a particular DMU under study (DMU_o). In the input-oriented CCR model as formulated in (1), the objective is to maximise the efficiency score of an DMU_o ($o = 1, \ldots, n$) under evaluation. u_j and v_k represent the jth input and the kth output weights for DMU_o.

$$\text{Max} \sum_{k=1}^{s} v_k y_{ko}$$

$$\text{s.t.} \sum_{j=1}^{m} u_j x_{jo} = 1$$

$$\sum_{k=1}^{s} v_k y_{ki} - \sum_{j=1}^{m} u_j x_{ji} \leq 0 \quad i = 1, \ldots, n$$

$$v_k, u_j \geq 0 \quad k = 1, \ldots, s, \quad j = 1, \ldots, m \tag{1}$$

The LP given above is solved for each DMU and the efficiency score (θ) of each DMU is obtained from each linear programme. A DMU is considered to be efficient if the optimal value for the LP problem is equal to one, otherwise it is inefficient.

The cross-efficiencies of *DMUs* can be found using the optimal input and output weights that other *DMUs* chose in model (1). However, the optimal weights obtained from the CCR model may not be unique (Baker and Talluri 1997) which deteriorates the effectiveness of the CE method in identifying good and poor performers. In order to overcome this limitation, Doyle and Green (1994) propose the following aggressive formulation:

$$\text{Min} \sum_{k=1}^{s} \left(v_k \sum_{i \neq o} y_{ki} \right)$$

$$\text{s.t.} \ \sum_{j=1}^{m} (u_j \sum_{i \neq o} x_{ji}) = 1$$

$$\sum_{k=1}^{s} v_k y_{ki} - \sum_{j=1}^{m} u_j x_{ji} \leq 0 \quad \forall \quad i \neq o$$

$$\sum_{k=1}^{s} v_k y_{ko} - \theta_o \sum_{j=1}^{m} u_j x_{jo} = 0$$

$$v_k, u_j \geq 0 \quad \forall k \text{ and } j \tag{2}$$

where DMU_o is the *DMU* under study, $\sum_{k=1}^{s} \left(v_k \sum_{i \neq o} y_{ki} \right)$ is the weighted output of a composite DMU, $\sum_{j=1}^{m} \left(u_j \sum_{i \neq o} x_{ji} \right)$ is the weighted input of the composite DMU and θ_o is the optimal efficiency of DMU_o derived from the CCR model. The formulation given in (2) is based on maximising the efficiency of the target DMU while minimising the efficiency of the composite DMU constructed from the other $n - 1$ DMUs. The cross efficiency of DMU_i, using the weights that DMU_o chose in the aggressive model is then

$$E_{oi} = \frac{\sum_{k=1}^{s} v_k y_{ki}}{\sum_{j=1}^{m} u_j x_{ji}} o, i = 1, 2, \ldots, n; E_i = \frac{1}{n} \sum_{o=1}^{n} E_{oi},$$

where v_k and u_j denote the optimal values obtained from model (2) and E_i is referred to as the cross efficiency score for DMU_i.