Loske, D., Mangiaracina, R., Regattieri, A., Klumpp, M. (2024):

„The Impact of Product Packaging Characteristics on Order Picking Performance in Grocery Retailing"

Journal of Business Logistics, Vol. 46, No. 1, p. e12400

Published Journal Article available at: https://doi.org/10.1111/jbl.12400

The Impact of Product Packaging Characteristics on Order Picking Performance in Grocery Retailing

Abstract

Increasing labor cost levels and workforce shortages have caused retailers to pay increased attention to their order-fulfillment operations, which continue to largely depend on manual order picking systems. The operations and logistics management literature suggests that optimizing tertiary packaging, which groups products into full unit loads for storage and shipping, is an important way to improve order picking performance. While most retailers handle products at the level of secondary packaging when fulfilling orders, this packaging level remains largely unexplored. To address this gap, we analyze 3,380,596 picks performed by 185 order pickers of 4,957 products in a grocery retail warehouse in Germany. Our findings indicate that secondary packaging characteristics directly affect order picking performance and that this effect is moderated by traditional product characteristics (e.g., product weight and volume), as well as elements of warehouse design (e.g., pick and stack levels). From a managerial perspective, our findings may help to bridge the gap between logistics managers and packaging engineers and provoke further research on the trade-off between operational and environmental performance.

Keywords: Packaging Logistics, Warehouse Performance, Order Picking, Retailing.

1 Introduction

Most retailers rely on manual picker-to-parts order picking, which is among the most cost-intensive warehousing activities, accounting for roughly 55% of total warehouse operating costs (De Koster et al. 2007, Boysen et al. 2021). The manual retrieval of products from storage locations in response to customer orders is extremely labor intensive. Increasing labor cost levels and labor shortages have thus intensified the pressure on warehouse managers (Vanheusden et al. 2022); in March 2023, the US Bureau of Labor Statistics reported a 2% increase in the Employment Cost Index for private industry compensation from its pre-pandemic level (U.S. Bureau of Labor Statistics 2023), whereas at the same time, as of July 2023, just 85% of job openings in the retail sector had been filled (U.S. Chamber of Commerce 2023). Against this backdrop, warehouse managers are increasingly interested in identifying factors affecting order picking performance to develop strategies to mitigate adverse effects while leveraging the factors that lead to positive outcomes.

Product packaging characteristics are among the factors most relevant to order picking performance and have received considerable attention in the operations and logistics management literature (Pålsson 2018). Neumann and Dul (2010) find that picking products from small plastic containers rather than pallets can reduce the physical workload of the order picker and increase productivity. Calzavara et al. (2017) and Hanson et al. (2018) also focus on packaging size (e.g., full- and half-size pallets), different-sized plastic containers, or cardboard boxes grouping components for warehouse storage in reserve areas and transit. While existing studies focus on how packaging characteristics impact order picking performance for tertiary packaging, there is a noticeable gap in the literature on the impact of secondary packaging characteristics. This gap is highly problematic for retail warehousing because changes in product packaging, such as reducing packaging material, inevitably impact secondary packaging, which is the level at which order pickers directly handle products during the picking process.

Companies are increasingly challenged with demands to play an active role in reducing the environmental impact of production and consumption impact and achieving environmental sustainability (Schoenherr et al. 2014). Product packaging is often considered a strategic response in sustainable business development (Kotzab et al. 2011). Recent changes in product packaging characteristics have been spurred by widespread public support for the United Nations (UN) Sustainable Development Goals (SDGs), see UN (2024). Manufacturers and retailers have continued to pledge their commitment to sustainability and to cutting back on their raw material use in furtherance of SDG 12, which is concerned with ensuring sustainable patterns of consumption and production (Gattiker et al. 2014, Lee and Murray 2019). Corporate sustainability goals are especially relevant for grocery retailers because food packaging still largely relies on non-returnable packaging, worsening consumers' material consumption footprint (Otto et al. 2021). Walmart offers an example of a sustainability pledge in its stated global goal of achieving 100% recyclable packaging and no unnecessary plastic packaging for its private brands by 2025 (Walmart 2024).

Given that product packaging has such a significant impact on order picking performance, and given the important role of order picking performance in warehouse operations, we address the gap in the research literature regarding secondary packaging noted above and pose the following research question: "How and to what extent do product packaging characteristics on the secondary level impact order picking performance?". The relevance of this research question is underlined by the ongoing shifts in secondary packaging as part of sustainability initiatives.

In this research, we collaborate with a German brick-and-mortar grocery retailer operating several warehouses for perishable and non-perishable products. To test our econometric model, we use a unique data set containing 3,380,596 picks performed by 185 order pickers for a product range of 4,957 dry food products in a single warehouse. We use the time elapsed in seconds between the completion of successive picks, pick $i-1$ and pick i, by picker j as the dependent variable in our multilevel regression model. We operationalize product-packaging characteristics by integrating the following independent variables: (1) whether the secondary packaging is in one piece, (2) whether the secondary packaging fully envelopes the primary packaging, and (3) the thickness of the cardboard box.

The remainder of this article is structured as follows. We discuss related literature on packaging logistics, order picking performance, humans in order picking, and the interplay between economic and environmental performance in Section 2. In Section 3, we describe the empirical setting and data set. In Section 4, we explain our model formulation, including all variables. Section 5 sets out our model-free results and the results of the multilevel regression. Because existing studies on sustainable supply chain management primarily focus on synergistic effects and overlooked trade-offs (Pagell and Shevchenko 2014), we include a discussion on the trade-off between economic and environmental performance in our specific context. In Section 7, we summarize the learnings and limitations of our study, as well as options for future research.

2 Related Literature

2.1 Packaging logistics

Our study builds on and contributes to three research streams: packaging logistics, order picking performance, and human workers in order picking systems. It further contributes to discussions on the interplay of economic and environmental performance. First, since we investigate the impact of product packaging characteristics, our study is naturally related to the sizable literature on packaging logistics. Product packaging is designed to maintain product quality throughout transport, storage, and manual handling along the entire supply chain (Hanson et al. 2018).

Product packaging systems have several levels, including primary, secondary, and tertiary packaging (Otto et al. 2021). Primary packages, also referred to as consumer units, are the smallest units for potential interaction and are in direct contact with the product. This level is especially relevant for shelf presentation and the communication power of product packaging systems (Underwood 2003). Secondary packaging groups a specified number of products into stock-keeping units, which are especially relevant for manual order picking processes in warehouses and during last mile transportation as well as in-store handling. Finally, tertiary packaging groups all stock-keeping units into full unit loads for warehouse storage in reserve areas and transportation (Pålsson 2018). Product packaging and logistics systems naturally interact in warehousing and transportation processes (Chan et al. 2006). Our main contribution to the packaging logistics literature is to provide empirical evidence of the direct impact of secondary-level product packaging characteristics on operational warehouse performance metrics.

2.2 Factors impacting order picking performance

Our research also contributes to the extensive literature on optimizing warehouse outcomes. Performance improvements in manual order picking systems have long been a popular area of research in operations and logistics management. Underperformance can result in high warehouse costs and dissatisfied customers (Gu et al. 2010). This interest has been extended in studies examining a variety of factors potentially impacting order picking performance along all sub-processes.

Following Tompkins et al. (2010), order picking involves the sub-processes of (1) traveling to pick locations, (2) reaching and bending to access pick locations, (3) physically retrieving products from storage locations, (4) documenting picking transactions, (5) sorting products into orders, (6) stacking products, and (7) searching for subsequent pick locations. Given the time these sub-processes consume in relation to total picking time, various studies examine manual product retrieval from storage locations.

Existing studies on sub-process (3) physically retrieving products from storage locations, the manual retrieval of products from storage locations, tend to focus on the design and evaluation of tertiary packages that group stock-keeping units into full unit loads for warehouse storage in reserve areas and for transport. Picking products from small plastic containers rather than pallets mitigates the physical demands on pickers while enhancing overall productivity, as demonstrated in Neumann and Dul (2010). Further, Calzavara et al. (2017) and Hanson et al. (2018) show that packaging variation – such as complete versus half-scale pallets, plastic containers of varying sizes, and the consolidation of components within cardboard boxes for warehouse storage and transportation – have a significant impact on order picking performance.

Some operations and logistics management studies also detail the role of tertiary packaging. Hanson and Finnsgård (2014) investigate the impact of unit load size on the efficiency of in-plant material supply for a Swedish automotive assembly firm in an empirical field-based study. Their findings indicate that the transition to smaller unit loads produced savings in the assembly process because the presentation of parts was improved. In an identical setting, Hanson et al. (2016) are more concerned with tilting unit loads and the position of products on a unit load. The authors find considerable differences between the front and rear of pallets and between their top and bottom sections and that picking time varies depending on each component's position in the container from which it is picked. Finnsgård and Wänström (2013) examine how component size, packaging type, and various aspects of component presentation impact order picking performance in a Swedish automotive company. They include packaging type and size for deep containers storing components as packaging variables and the size and weight of the component to be picked as part-property variables. They report that packaging, angle, and picking height have the greatest impact on order picking performance. Unfortunately, they neglect the characteristics of the components that need to be picked (equal to our secondary packages) from deep containers (equal to our tertiary packages).

In summary, prior work accounts for tertiary packaging, grouping products into full unit loads for warehouse storage in reserve areas and for transit, but not for the characteristics of secondary packaging. We thus contribute empirical evidence to the literature on factors impacting order picking performance in a new setting, one in which order pickers process products at the level of secondary packaging. This is especially relevant for brick-and-mortar grocery retailers, as they fulfill most store orders at this level.

2.3 Human workers in order picking systems

Prior studies in operations and logistics management have long recognized that worker behavior is highly heterogeneous, leading to between- and within-worker differences in performance (Matusiak et al. 2017). Understanding the cause of such variations is highly relevant for scholars and practitioners since these directly impact metrics of operational performance metrics, such as individual order picking performance (Sun et al. 2022). Fluctuations in the performance of individual workers and between workers are particularly important in manual picker-to-parts order picking systems, where human involvement is a defining feature of the order fulfillment process (Grosse and Glock 2015, De Vries et al. 2016).

Various studies have considered how best to manage heterogeneity from perceptual, mental, physical, and psycho-social human factors to improve performance (Grosse et al. 2017, Corbett 2023). We expand this research by examining how the characteristics of secondary-level product packaging impact warehouse design (e.g., pick and stack level) and product-related factors (e.g., weight and volume). We contribute to the operations and logistics management literature with suggestive evidence of how to mitigate the adverse performance effects of packaging characteristics by these various factors into operational and tactical decision models.

2.4 The interplay of operational and environmental performance

Balancing economic and environmental performance remains a considerable challenge for companies (Schoenherr et al. 2014, Davis-Sramek et al. 2018). Sheffi (2018) argues that while "some environmental initiatives

[...] support the mission of the business, such as energy savings that also reduce costs, a business should not go too far". This statement points to the potential interaction between economic and environmental performance and its potential to produce trade-off, lose-lose, or win-win situations (Govindan et al. 2020). We explore this further and enlarge our understanding of the economic dimensions that we relate to 'operational' performance; that is, economic value perspectives, like cost, are included as operational but so are more process and competitive dimensions, such as speed and time to market.

Existing studies on the interplay of operational and environmental performance evaluate the economic performance of firms using market and accounting-based measures and operational perspectives (Zhu and Sarkis 2004). The environmental performance of firms includes activities that reduce their environmental impact and is an element of the triple bottom line approach for sustainable supply chain management; Bentahar and Benzidia (2018) defines this approach as "(...) the integration of economic, environmental and social dimensions into the management of intra- and inter-organizational flows (...) with the objective of creating sustainable value".

We conceptualize this interplay of operational and environmental performance in a theoretical framework (Hahn et al. 2010). The matrix in Figure 1 has operational and environmental performance on the two axes, following Figge and Hahn (2012). Any improvements in operational performance are captured in Fields 1 and 4, while improvements in environmental performance are captured in Fields 3 and 4. Field 2 thus represents undesirable outcomes with regard to operational and environmental performance, and Field 4 contains a win-win situation.

These combinations are relevant for corporate decision-making regarding outcomes to seek or avoid. However, these obviously do not require dedicated analysis or research insight other than clear criteria and accurate analysis of data acquisition (to ensure that the relevant decision is placed exactly in the relevant field). The more interesting interplay also leads to two specific trade-off situations in which companies target operational performance at the cost of imposing an environmental burden, as shown in Field 1 (Type-A trade-off); alternatively, Field 3 depicts a Type-B trade-off with positive environmental performance implications but negative impacts on operational performance.

Specific (market, accounting, and operational) perspectives on economic and environmental performance are identified in the literature, and studies identify trade-offs, win-win situations, undesirable outcomes, and trade-offs developing into win-win situations (Erbetta et al. 2023, Ünal and Sinha 2023). Mao et al. (2017) examine the interplay between reduced carbon emissions and firm performance, finding a Type-B trade-off in which reducing emissions through process improvements improves environmental performance but can negatively impact financial performance. Another Type-B trade-off is identified by Mahapatra et al. (2021) who examine the impact of internal and external initiatives on firms' carbon footprints and find that most firms pursuing carbon reduction to increase environmental performance were still awaiting a positive impact on operational performance. Jabbour et al. (2014) find that green strategies, for example, reducing materials or pollution, can lead to increased process efficiency, a win-win situation in Figure 1.

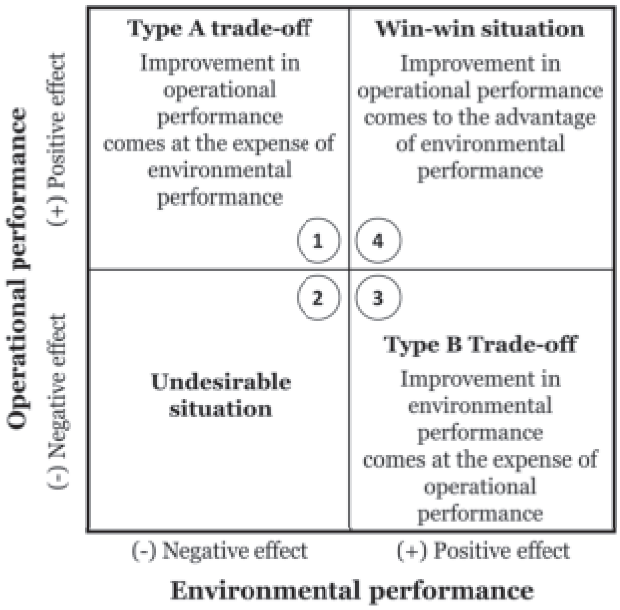

Figure 1: Framework for the interplay of economic and environmental performance following Figge and Hahn (2012).

Kumar and Rodrigues (2020) report the possibility of increasing environmental *and* operational performance; for example, eco-friendly products are cheaper to produce through integrated lean and green practices in design and service delivery (a win-win situation). Esfahbodi et al. (2023) provide an example of a transition from a trade-off into a win-win situation, finding that green supply chain management yields trade-offs between operational and environmental performance. Finally, the trade-off issue is connected to the potential input of minimum requirement definitions in both areas (Seuring and Müller 2008) and possible rebound issues (Chenavaz et al. 2021).

Our contribution to the operational and environmental performance literature is an operational perspective on trade-offs and win-win situations in warehouse order picking. We then test moderation effects, which allows us to examine whether there are transitions from trade-offs to win-win situations. Operational performance, as a form of economic performance, increases when the time needed for order picking decreases (Matusiak et al. 2017, Batt and Gallino 2019) and environmental performance increases when emissions or resource use decrease (Golicic and Smith 2013). We refer to this form of economic performance as operational performance.

We contrast the potential interplay in the following example. Consider the reduction of cardboard in secondary product packaging. Reducing 1 kilogram of cardboard for product packaging results in a 0.94-kilogram decrease in carbon emissions (Brogaard et al. 2014), improving environmental performance. However, this reduction can have varying effects on operational performance: Thinner cardboard boxes might reduce product stability, making packages harder to pick and increasing the time needed for order picking (Hanson et al. 2018). In this case, environmental performance improvements come at the expense of operational performance, a Type-B trade-off.

In contrast, thinner cardboard for boxes could also reduce weight, making packages easier to pick and decreasing the time needed for order picking (Finnsgård and Wänström 2013). An improvement in operational performance is, thus, beneficial for environmental performance, and is a win-win situation. In this regard, and drawing on the empirical results in Sections 4 and 5, we return to discussing such trade-offs, win-win situations, and possible transition pathways in Section 6.

3 Empirical Setting and Data Description

This empirical secondary-data study was conducted in collaboration with a German brick-and-mortar retailer operating several warehouses for perishable and non-perishable products delivered to more than 7,000 stores nationally. In the warehouse that is the focus here, cooled and non-cooled warehouse sections are operated using picker-to-parts order picking with vehicle support by industrial trucks. The specific warehouse under examination, at the time of analysis, stored 4,957 dry food products in 49 aisles organized in 10 picking zones. Storage is located on the ground level, and picker travel is organized in a U-shaped pattern.

The average number of picks per tour for order pickers, including multiple visits to storage locations, is 106.42, with an average of 243.06 storage-location visits per hour. Human forklift operators supply the ground-level picking zones with unit loads stored in reserve areas in the upper levels of the high-rise shelves. Figure 2 depicts the picker-to-parts order picking system studied.

Our study focuses on non-cooled products, allowing us to control for product-specific effects on order picking performance. Additionally, the purchasing and logistics departments work independently to negotiate prices with suppliers. Therefore, we can control for potential confounding factors, for example, economically driven decisions on the design of the product packaging system that originate from the logistics department and could impact the dependent variable, order picking performance. Because all products are controlled for quantity, quality, and packaging in the incoming goods department, we can exclude article numbers with respect to which packaging changes occurred during the period investigated. This is especially relevant for promotional products.

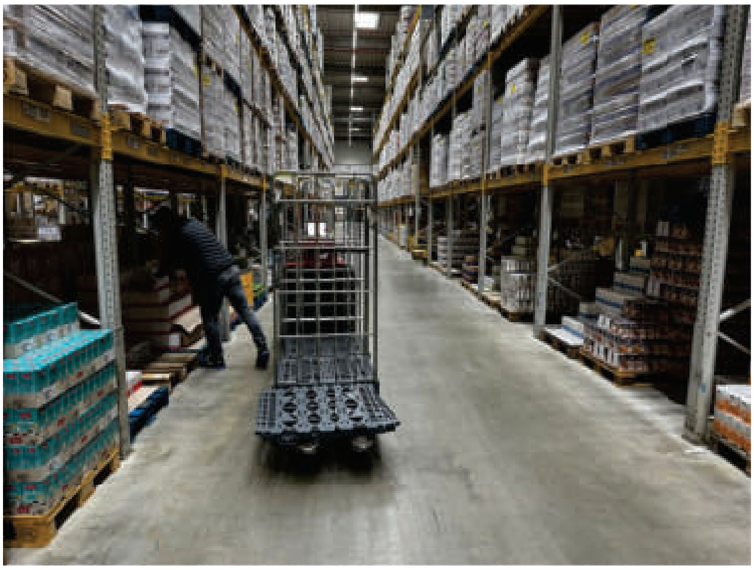

Figure 2: Picker-to-parts order picking system with vehicle support by industrial trucks.

We analyze order picking data between June and December 2021. Most warehouse management system (WMS) applications store extensive, detailed log data on order picking processes. We utilize such log datasets to construct a model capable of evaluating the accelerating and decelerating variable impacts, operationalizing various aspects of packaging. Our data set, extracted from the company's WMS, includes data on batch identification (ID), pick ID, picker ID, load unit ID, article number, number of units picked, length, width, and height of secondary product packaging, volume of secondary product packaging, weight of the product and secondary product packaging, timestamp for each pick, and slot address per pick.

Our initial data set includes 4,349,262 picks performed by 216 order pickers. Because we use real-world data, the log data are polluted. There are several reasons for this, including personnel breaks or system breakdowns. Therefore, we exclude all picks lasting longer than 100 seconds, as they have been identified as invalid data in the underlying scenario (dropping 146,562 observation points). Next, we control the speed of the industrial trucks and exclude all picks with an average travel speed higher than 3.33 m/s, (dropping 54,339 observation points). We also exclude all batches for which the load unit used for picking is not a standardized roll cage, (dropping 647,311 observation points). Finally, we exclude all packaging changes during the investigation period, (dropping 120,454 observation points). After cross-validating all data-cleaning rules with the company, our final data set comprises 3,380,596 picks performed by 185 order pickers for a product range of 4,957 non-cooled products. We describe the model applied to this data set in the next section. Table 1 summarizes the descriptive statistics for all variables.

Table 1: Descriptive statistics.

No.	Variable	Description of operationalization	Operationalization	Mean	SD.
1	Pick_Time	Timestamps for the beginning and the end of the picking process are used to set the border of the total event time.	Continuous	20.88	21.27
2	One_Piece_Secondary_Packaging	The secondary packaging is one-piece packaging grouping primary packages into one SKU.	Binary dummy: 1 = Yes (67,18 %), 0 = No (32.82 %)	0.82	0.39
3	Fully_Enveloping_Secondary_Packaging	The secondary packaging fully envelops the primary packages, grouping them into one SKU.	Binary dummy: 1 = Yes (80,48 %), 0 = No (19.52 %)	0.70	0.46
4	Carton_Box_Thickness	Whenever the secondary package is a carton, the thickness in millimeters is used as a proxy for package stability.	Continuous	2.49	1.21
5	Product_Weight	Weight in kilograms per SKU, including the product, primary packaging, and secondary packaging.	Continuous	5.13	3.25
6	Picker_Experience	cumulative number of picks per order picker and in the dataset.	Continuous	12,252.47	8.814,77
7	Travel_Distance	The distance in meters from pick location to pick location.	Continuous	15.25	27.98
8	Pick_Level	The pick level of the picking location.	Binary dummy: 1 = chest level (50,77%), 0 = ground level (49,23%)	0.54	0.61
9	Picks	Number of picks from one picking location.	Continuous	1.33	1.35
10	Stack_Level	One batch consists of several picks. This variable quantifies the number of products stacked on a unit load and serves as a proxy for stack level.	Continuous	71.53	59.15
11	Product_Volume	Volume of the secondary package in liters.	Continuous	12.91	12.17
12	Primary_Packaging	Number of primary packages combined into a single secondary package representing one SKU.	Continuous	10.46	10.21

Notes: SD = Standard Deviation. Stock keeping unit (SKU) does not refer to a single consumer unit. Rather, it encompasses several consumer units, and we control for the number of consumer units by the primary packages variable (Number of primary packages packed into one secondary package representing one SKU.)

4 Model Formulation

Our empirical analysis focuses on estimating the impact of product packaging characteristics on order picking performance , such as (1) one-piece secondary packaging, (2) fully enveloping secondary packaging, and (3) the thickness of the carton. Since our econometric analysis in Section 5 presents variations of a single econometric model, we first describe our base model before we proceed with the extended models including interaction terms. Due to our longitudinal research design, we measure order picking performance for each order picker repeatedly over time. Because repetitive measurement of individuals violates the assumption of independence in linear regression models, we propose a multilevel model, also termed a hierarchical or mixed-effects model. This allows us to measure individual order pickers more than once without artificially inflating our estimates.

Our dependent variable of interest is order picking performance: $Pick\_Time_{ijk}$ is defined as the time elapsed in seconds between the completion of successive picks (pick $i-1$ and pick i) by picker j in aisle k. The clock starts when the order picker confirms the start of a pick by pushing "next" on a personal digital assistant mounted on the accompanying industrial truck. The device maintains a constant wireless connection with the WMS, documenting relevant time stamps. After traveling to the pick location and picking the required number of products, the clock stops when the fulfillment of these process steps is confirmed by the picker pushing a symbol representing a unit load on the personal digital assistant. $PICK\_TIME_{ijk}$ is operationalized as a continuous metric variable and is frequently used in logistics and operations management research to evaluate performance in order picking systems (Matusiak et al. 2017, Batt and Gallino 2019).

After defining our dependent variable, we specify fixed and random effects in the multilevel model; for this purpose, we calculate the intraclass correlation coefficient (ICC) for two separate no-predictor models and test whether our data set reflects within- and between-group variance of (1) order pickers and (2) aisles. The ICC values of the no-predictor models can be interpreted as the total variance in the dependent variable, $Pick\_Time_{ijk}$, originating from picker- or aisle-related variation. This variance is attributable to between-picker (between-aisle) rather than within-picker (within-aisle) variation over time. Higher values also indicate a non-trivial degree of observation non-independence, indicating traditional regression approaches may be inappropriate. The ICC value for $Pick\_Time_{ijk}$ is 19% for pickers (9% for aisles), meaning that approximately one-fifth of the variance is attributable to between-picker differences (a tenth attributable to between-aisle differences), the remainder is explained by within-picker (within-aisle) variability over time. Hence, $Pick\_Time_{ijk}$ differs between order pickers j and aisles k, suggesting that estimating more complex models with hierarchies and temporal change is warranted.

We propose a mixed-effects model with random intercepts because we find relevant within- and between-group variance in order pickers and aisles. We allow order picker j to vary as a random main intercept α_j to account for individual differences. Further, we allow aisle k to vary as a random intercept δ_j to account for aisles-related differences (e.g., products with particular characteristics stored in specific aisles).

Finally, $(\alpha\delta)_{jk}$ denotes the random interaction term for the relationship between picker j and aisle k. Note that $(\alpha\delta)_{jk}$ is treated as a crossed and not a nested random effect because there is no fixed assignment of order picker j to aisle k; Matusiak et al. (2017) report an ICC of 10.3% of the total variance in the natural logarithm of time and 13.1% of the non-transformed time for quantifying differences between pickers.

We examine three key independent variables. For the first, *One_Piece_Secondary_Packaging*, we screen 4,957 products manually through primary data collection and capture data on the secondary packing by product identification number. This number is then used to merge our primary data with the WMS data set. We code *One_Piece_Secondary_Packaging* as a binary dummy variable (0 = not one-piece secondary packaging, 1 = one-piece secondary packaging). The second is *Fully_Enveloping_Secondary_Packaging*, using the same procedure employed with one-piece packaging, and we code *Fully_Enveloping_Secondary_Packaging* as a binary dummy variable (0 = no fully-enveloping secondary packaging, 1 = fully enveloping secondary packaging). The third, *Carton_Box_Thickness*,is the caliper-measured thickness of the carton box (in millimeters); our variable for the carton thickness is thus continuous. Figure 3 depicts two of the independent variables for ease of understanding the independent variables.

Figure 3: Exemplary visualization of product packaging characteristics.

We then integrate several control variables that we consider essential for the validity of our study and support their inclusion with theoretical boundaries, their relation to the independent variables, and previous research postulating empirical relationships between the particular control and variables in our study (Cuervo-Cazurra et al. 2016). We determine which control variables to include in our econometric model by following the decision tree proposed by Bernerth and Aguinis (2016).

First, we control for $Product\_Weight$, $Product\_Volume$, and $Primary\_Packaging$, which are identified as relevant product characteristics in empirical research (Wänström and Medbo 2008, Chabot et al. 2017) and might also interact with other packaging characteristics that we treat as independent variables. We also add the $Pick\_Level$ and $Stack\_Level$ to account for the height of picking and stacking. Petersen et al. (2005) include height in their empirical models. Picking height might interact with product packaging characteristics because it reflects the physical effort required to retrieve and stack different products (Finnsgård and Wänström 2013, Hanson and Finnsgård 2014, Hanson et al. 2018).

Studies on order picking clearly establish that travel distance impacts order picking performance (Pan et al. 2014, Masae et al. 2020). We need to control for travel distance because the time required for picker travel is included in our dependent variable. Batt and Gallino (2019) integrate the number of picks per storage location as pick density and find significant interactions with product characteristics like color. Therefore, we control for the number of picks retrieved from a storage location with $Picks$. Finally, the literature on human factors shows that cumulative experience impacts order picking performance (Batt and Gallino 2019, Loske 2022). Because this is especially true for manual material handling, we control for human experience effects with $Picker\_Experience$. We then add a final variable τ, capturing time-related effects as a control for month, day of the week, and hour of the day; ϵ_j is the error term. Our base model is denoted as follows:

$$\begin{aligned} Pick\_Time_{ijk} =& \alpha_{0j} + \delta_{0k} + (\alpha\delta)_{0jk} + \beta_{1j} One\_Piece\_Secondary\_Packaging_i + \\ & \beta_{2j} Fully\_Enveloping\_Secondary\_Packaging_i + \beta_{3j} Carton\_Box\_Thickness_i + \\ & \beta_{4j} Product\_Weight_i + \beta_{5j} Picker\_Experience_i + \beta_{6j} Travel\_Distance_i + \\ & \beta_{7j} Pick\_Level_i + \beta_{8j} Picks_i + \beta_{9j} Stack\_Level_i + \\ & \beta_{10j} Product\_Volume_i + \beta_{11j} Primary\_Packaging_i + \tau + \epsilon_{ijk} \end{aligned} \tag{1}$$

$$\alpha_{0j} = \gamma_{00} + \upsilon_{0j} \tag{2}$$

$$\delta_{0k} = \gamma_{00} + \upsilon_{0k} \tag{3}$$

$$(\alpha\delta)_{0jk} = \gamma_{00} + \upsilon_{0jk}, \tag{4}$$

where

$$One\_Piece\_Secondary\_Packaging_i = \begin{cases} 1, & \text{if one-piece secondary packing} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$Fully\_Enveloping\_Secondary\_Packaging_i = \begin{cases} 1, & \text{if fully enveloping secondary packaging} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Multilevel samples should have a minimum of 30 observations at each level to ensure statistical power (McCoach and Black 2012). In our analysis, Level 1 includes picks in the full sample (N = 3,380,596) nested within order pickers (Level 2, N = 185. We employ the lme4 (Bates 2022) and multilevel (Bliese 2022) packages in R for Windows for analysis. To avoid possible concerns about the validity of our econometric model, we test correlation for all variables and attach a correlation table in Appendix 1. Further, we calculate each variable's variance inflation factors (VIFs). The highest VIF is 10.81 for $Product\_Volume$, indicating that cross-correlation effects, which could cause artificial inflation of estimators and p-values, are not a serious issue in our model.

5 Empirical Results

5.1 Model-free evidence

We start by examining how established product characteristics like weight and our packaging characteristics impact the dependent variable $Pick\_Time$. For this purpose, we visualize model-free evidence, as recently proposed by Davis-Sramek et al. (2023). The basic idea is to facilitate understanding of our results without the need to understand all the conditions of our multilevel model. Figure 4 is a visualization of $One\_Piece\_Secondary\_Packaging$ and $Fully\_Enveloping\_Secondary\_Packaging$ on the x-axis and the $Pick\_Time$ on the y-axis (in seconds). Additionally, we build three diagrams according to product weight: with a standard deviation (SD) of -1 and +1 to the left and right, respectively, and the mean in the middle. As a fourth dimension, we visualize three lines depending on $Carton\_Box\_Thickness$, again differentiated by SD and mean.

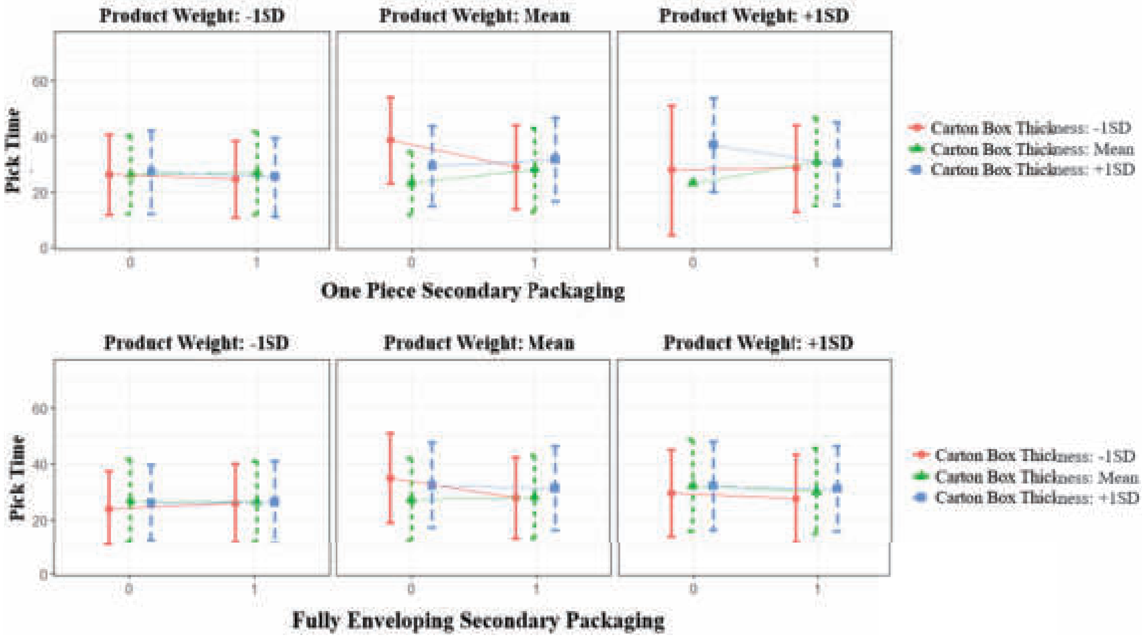

Figure 4: Model-free evidence for one-piece and fully enveloping secondary packaging.

We find different means (red circle, green triangle, blue square) and SDs (whiskers of second quartile below and third quartile above the mean) for different product weights. However, the relationship is not linear, as proposed in previous studies (e.g., higher $Product\_Weight$ increases $Pick\_Time$). Instead, we find that for each $Product\_Weight$ cluster, $One\_Piece\_Secondary\_Packaging$ and $Fully\_Enveloping\_Secondary\_\ Packaging$ produce variability in $Pick\_Time$, which becomes even stronger when the fourth variable $Carton\_Box\_Thickness$ is included. The two most important main results are summarized below.

First, $Product\_Weight$ impacts $Pick\_Time$ but not necessarily in a linear fashion. Further, there are relevant interactions of product characteristics, such as weight and our selected packaging characteristics, that demand a more detailed analysis in regression-based models.

Second, we are interested in the batch and packaging characteristics impacting the dependent variable, $Pick\_Time$. We thus differentiate between low- and high-frequency picks. Figure 5 depicts One_Piece $Secondary\_Packaging$ and $Fully\_Enveloping\_Secondary\_Packaging$ on the x-axis and $Pick\_Time$ in seconds on the y-axis. $Carton\_Box\_Thickness$ is integrated as a fourth variable, as we did for the previous model-free visualization. Here, we identify minimal variation caused by $Carton\_Box\_Thickness$. The variability mainly originates in $One\_Piece\_Secondary\_Packaging$ and $Fully\_Enveloping\_Secondary\_Packaging$ but is less obvious than the variability related to $Product\_Weight$ visualized in Figure 4.

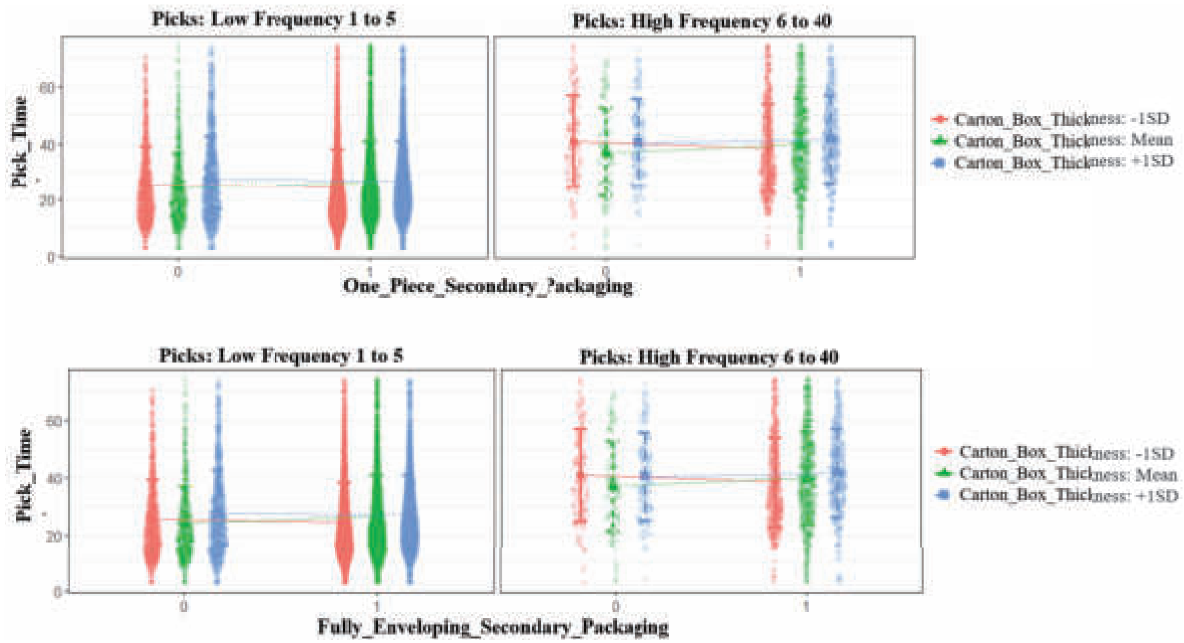

Figure 5: Model-free evidence for pick frequency.

5.2 Main effects: Product packaging characteristics

We now examine the main impact of product packaging characteristics on $Pick\_Time$. Note that a positive coefficient is related to a higher predicted $Pick\_Time$ and, therefore, a negative impact on order picking performance; a negative coefficient is related to a lower predicted $Pick\_Time$, with a positive effect on order picking performance. We draw the readers' attention to Table 2 for Models (1) and (2). In Model (1), we integrate all control variables and add the independent variables of interest in Model (2). The Lower Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values and the higher r-squared values indicate an improved model fit and higher explained variance when product packaging characteristics are included in the model. This represents a first important finding: Neglecting product packaging characteristics inflates the strength of the control variables and leads to incorrect assessments.

Table 2: Muli-level model: Main effects.

	Dependent variable: Log.PickTime(inSec.)	
	(Model 1)	(Model 2)
Independent variables of interest		
One Piece Secondary Packaging		0.0158 ($p < 0.001$)
Fully enveloping secondary packaging		−0.0283 ($p < 0.001$)
Carton box thickness		−0.0016 ($p = 0.0042$)
Control variables		
Product weight	0.0167 ($p < 0.001$)	0.0111 ($p < 0.001$))
Product volume	−0.0017 ($p < 0.001$)	0.0033 ($p < 0.001$)
Primary packaging	0.0021 ($p < 0.001$)	0.0009 ($p < 0.001$)
Stack level	−0.00002 ($p = 0.1727$)	−0.0010 ($p < 0.001$)
Pick level	0.0074 ($p < 0.001$)	0.0114 ($p < 0.001$)
Travel distance	0.0090 ($p < 0.001$)	0.0091 ($p < 0.001$)
Picks	0.1170 ($p < 0.001$)	0.1170 ($p < 0.001$)
Picker experience	−0.0005 ($p = 0.0135$)	−0.0014 ($p < 0.001$)
Time-fixed effect	Included	Included
Observations	3,380,591	2,523,254
Unique products	4,957	4,957
Pickers	185	185
AIC	9,928,794	7,312,882
BIC	9,928,951	7,313,098
R Squared	0.099	0.132

We find evidence that $Fully\_Enveloping\_Secondary\_Packaging$ negatively and significantly impacts $Pick\_Time$. When secondary packaging fully envelopes the primary product packaging, $Pick\_Time$ is reduced by 2.83%. Additionally, we find that $Carton\_Box\_Thickness$ can decrease $Pick\_Time$ by .16% for each millimeter of thickness. While a potential improvement of 4.43% (2.83% + 1.6%) is hard to capture, we draw on our example of 185 order pickers working approximately 7.75 hours per day. Priced using the average European minimum wage, which ranges from 2.00 euro in Bulgaria to 13.05 euro in Luxembourg (Statista 2022), the potential ranges from 688.200 euro to 4.490.505 euro per year (assuming 240 working days per year).

Finally, we report a counter-intuitive main effect for $One\_Piece\_Secondary$ $\_Packaging$, which increases $Pick\_Time$ by 1.58% and, therefore, has a negative impact on order picking performance. Table 2 summarizes the findings of our multilevel model. Note that we scale picker experience by 1/1,000 to improve the transparency of reported results.

5.3 Moderating effects: The role of product weight and volume

In addition to our base model, we estimate an extended model in which we add $Product\_Weight$ and $Product\_Volume$ as potential moderators for product packaging characteristics impacting $Pick\_Time$:

$$\begin{aligned}
Pick_Time_{ijk} =& \alpha_{0j} + \delta_{0k} + (\alpha\delta)_{0jk} + \beta_{1j} One_Piece_Secondary_Packaging_i + \\
& \beta_{2j} Fully_Enveloping_Secondary_Packaging_i + \beta_{3j} Carton_Box_Thickness_i + \\
& \beta_{4j} One_Piece_Secondary_Packaging_i \times Product_Weight_i + \\
& \beta_{5j} Fully_Enveloping_Secondary_Packaging_i \times Product_Weight_i + \\
& \beta_{6j} Carton_Box_Thickness_i \times Product_Weight_i \\
& \beta_{7j} One_Piece_Secondary_Packaging_i \times Product_Volume_i + \\
& \beta_{8j} Fully_Enveloping_Secondary_Packaging_i \times Product_Volume_i + \\
& \beta_{9j} Carton_Box_Thickness_i \times Product_Volume_i + \\
& \beta_{10j} One_Piece_Secondary_Packaging_i \times Stack_Level_i + \\
& \beta_{11j} Fully_Enveloping_Secondary_Packaging_i \times Stack_Level_i + \\
& \beta_{12j} Carton_Box_Thickness_i \times Stack_Level_i + \\
& \beta_{13j} One_Piece_Secondary_Packaging_i \times Pick_Level_i + \\
& \beta_{14j} Fully_Enveloping_Secondary_Packaging_i \times Pick_Level_i + \\
& \beta_{15j} Carton_Box_Thickness_i \times Pick_Level_i + \beta_{nj} Controls_i + \tau + \epsilon_{ijk}
\end{aligned} \tag{7}$$

$$\alpha_{0j} = \gamma_{00} + \upsilon_{0j} \tag{8}$$

$$\delta_{0k} = \gamma_{00} + \upsilon_{0k} \tag{9}$$

$$(\alpha\delta)_{0jk} = \gamma_{00} + \upsilon_{0jk} \tag{10}$$

where,

$$One\_Piece\_Secondary\_Packaging_i = \begin{cases} 1, & \text{if one-piece secondary packing} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$Fully\_Enveloping\_Secondary\_Packaging_i = \begin{cases} 1, & \text{if fully enveloping secondary packaging} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Table 3 summarizes the results, with separate interaction effects in Models (1) and (2) as well as in the joint model of Formula (7) in Model (5). Additionally, we select two interaction effects for visualization through interaction plots in Figures 6 and 7.

Figure 6: Interaction Plot: Product weight and one-piece secondary packaging.

Regarding the interaction of product weight and product packaging characteristics visualized in Figure 6, we find that $Product\_Weight$ generally increases $Pick\_Time$. However, the impact of $Product\_Weight$ on $Pick\_Time$ is significantly weaker for $One\_Piece\_Secondary\_Packaging$, starting at a cross-over point of 4.4 kilograms $Product\_Weight$ (the cross-over point in the interaction plot, Figure 6). Hence, the impact of $One\_Piece\_Secondary\_Packaging$ depends on a threshold value for $Product\_Weight$. We also find that the thickness of a carton can weaken the effect of $Product\_Weight$ on $Pick\_Time$. Therefore, the thicker the carton, the lower the effect of $Product\_Weight$ on $Pick\_Time$. This is depicted visually in Figure 7.

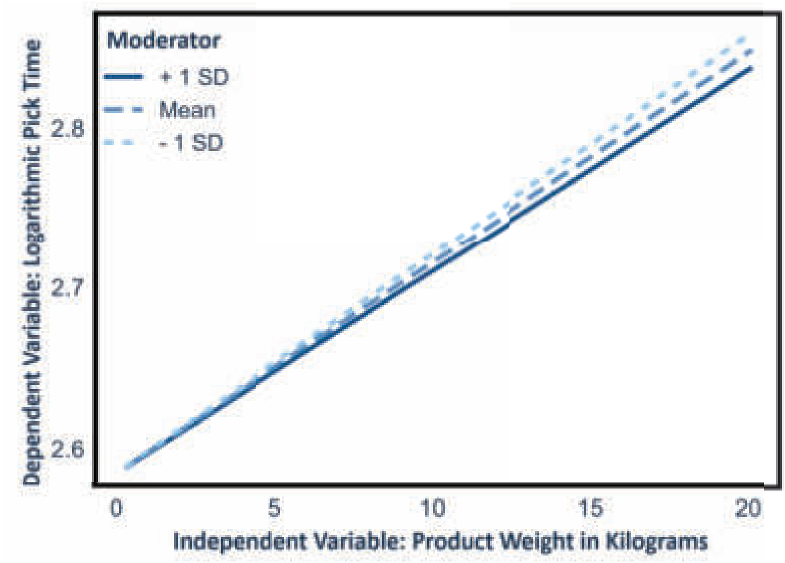

Figure 7: Interaction plot: Product weight and carton thickness.

Table 3 summarizes all results for the interaction effects. When comparing Models (3) and (4) (with one interaction effect per model) and Model (7) (as a full model with all interaction effects and controls), we find the latter offers the best fit, indicated by having the lowest AIC and BIC values. Therefore, it seems necessary to consider the interactions between product packaging and product characteristics, as well as storage and retrieval height. The main effects for Model (7) are that $Fully\_Enveloping\_Secondary\_Packaging$ (6.45% reduction) and $One\_Piece\_Secondary\_Packaging$ (19.44% reduction) have a negative and significant impact on $Pick\_Time$. As highlighted in Figure 6, this is especially true for products weighing at least 4.4 kilograms. $Carton\_Box\_Thickness$ has a positive and significant impact on $Pick\_Time$. Each additional millimeter of thickness increases $Pick\_Time$ by 0.67% for the main effect. However, for products that are heavy, thick cartons can decrease $Pick\_Time$, as indicated by the interaction term $Carton\_Box\_Thickness \times Product\_Weigth$.

Table 3: Multilevel model: Interaction effects.

	Dependent variable: Log.PickTime(inSec.)				
	(Model 3)	(Model 4)	(Model 5)	(Model 6)	(Model 7)
Independent variables					
One-piece secondary packaging	0.0469	−0.0194	−0.0064	−0.0351	−0.0645
	$p < 0.001$	$p < 0.001$	$p = 0.0441$	$p < 0.001$	$p < 0.001$
Fully enveloping secondary packaging	−0.1203	−0.0829	0.0044	−0.0399	−0.1944
	$p < 0.001$	$p < 0.001$	$p = 0.2054$	$p < 0.001$	$p < 0.001$
Carton thickness	0.0020	0.0050	−0.0015	−0.0024	0.0067
	$p = 0.0437$	$p < 0.00100$	$p = 0.1145$	$p = 0.0385$	$p = 0.0006$
Moderators					
Product weight	0.0103	0.0112	0.0115	0.0111	0.0075
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Product volume	0.0031	−0.0009	0.0033	0.0033	0.0023
	$p < 0.001$	$p = 0.0139$	$p < 0.001$	$p < 0.001$	$p < 0.00100$
Stack level	−0.0010	−0.0010	−0.0007	−0.0010	−0.0010
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Pick level	0.0170	0.0119	0.0115	−0.0121	−0.0510
	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.0214$	$p < 0.001$
Interaction terms					
One-Piece Secondary Packaging × Product Weight	−0.0101				−0.0107
	$p < 0.001$				$p < 0.001$
Fully enveloping secondary packaging × Product Weight	0.0166				0.0173
	$p < 0.001$				$p < 0.001$
Carton thickness × Product Weight	−0.0009				−0.0004
	$p < 0.00103$				$p = 0.0591$
One-piece secondary packaging × Product Volume		0.0028			0.0038
		$p < 0.001$			$p < 0.001$
Fully enveloping secondary packaging × Product Volume		0.0049			0.0008
		$p < 0.001$			$p = 0.0176$
Carton thickness × Product Volume		−0.0005			−0.0005
		$p < 0.001$			$p < 0.001$
One-piece secondary packaging × Stack Level			0.0002		0.0001
			$p < 0.001$		$p = 0.0002$
Fully enveloping secondary packaging × Stack Level			−0.0005		−0.0002
			$p < 0.001$		$p < 0.0012$
Carton thickness × Stack Level			−0.000001		0.00002
			$p = 0.8953$		$p = 0.0368$
One-piece secondary packaging × Pick Level				0.0362	0.0390
				$p < 0.001$	$p < 0.001$
Fully enveloping secondary packaging × Pick Level				0.0056	0.0552
				$p = 0.2363$	$p < 0.001$
Carton thickness × Pick Level				0.0003	−0.0016
				$p = 0.7106$	$p = 0.0582$
Controls	Included	Included	Included	Included	Included
Observations	2,523,254	2,523,254	2,523,254	2,523,254	2,523,254
AIC	7,311,214	7,312,410	7,312,625	7,312,613	7,310,417
BIC	7,311,469	7,312,664	7,312,880	7,312,868	7,310,774

6 Discussion

Drawing on the discourse in the preceding section on evaluating operational performance as a form of economic performance, we now elaborate on potential trade-offs and win-win and undesirable situations. We derive a negative effect on operational performance whenever the coefficients for the main effects of product packaging on pick time are positive and thus increase the time required for order picking. In contrast, we derive a positive effect on operational performance whenever the coefficients for the main effects of product packaging on pick time are negative and thus decrease the time and cost of order picking. We find a negative effect for fully enveloping secondary packaging (a positive effect on operational performance) compared to the non-fully enveloping secondary packaging (a negative effect on operational performance).

We derive the environmental dimension by again comparing fully and non-fully enveloping secondary packaging. The latter uses less packaging material than the former. Drawing on the argument of relative performance effects discussed regarding the multilevel regression, we find that non-fully enveloping secondary packaging positively affects environmental performance, and fully enveloping secondary packaging has a negative effect. The best option would be to calculate the exact reduction of carbon emissions with the data in Brogaard et al. (2014). However, at this point, we cannot separate our product weight data into product and packaging weight.

After defining and evaluating the operational and environmental performance effects of fully and non-fully enveloping secondary packaging, we can establish their position in the proposed theoretical framework. With a positive effect on operational performance but a negative effect on environmental performance, fully enveloping secondary packaging presents a Type-A trade-off. In this case, improvements in operational performance due to reduced order picking costs come at the expense of reduced environmental performance due to the carbon emissions entailed in increased packaging. By contrast, non-fully enveloping secondary packaging is defined as a Type-B trade-off due to its positive effect on environmental performance (lower resource use and carbon emissions due to less packaging) but a negative effect on operational performance due to the increased duration and cost of order picking. The fully filled circles in Figure 8 show the position of fully and non-fully enveloping secondary packaging in the proposed theoretical framework.

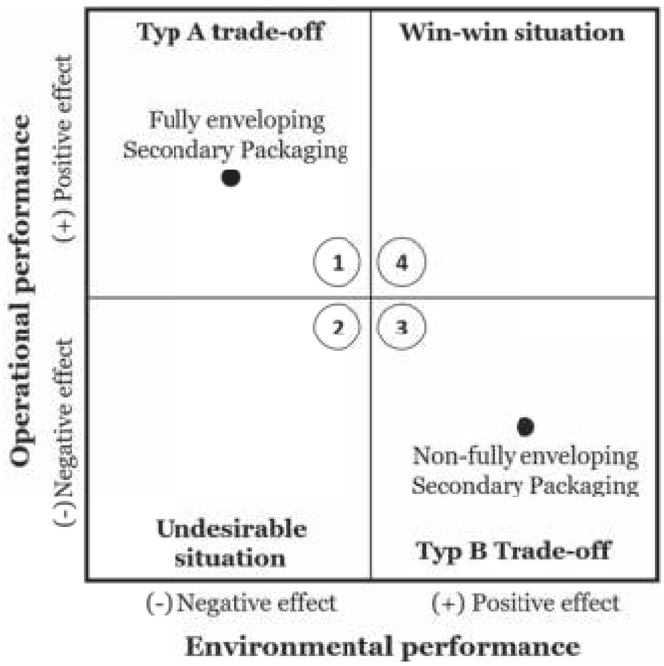

Figure 8: Trade-off positions of fully and non-fully enveloping secondary packaging.

Based on our interaction terms, we can furthermore determine whether there is a change in the positions of fully and non-fully enveloping secondary packaging in the theoretical framework when there is a change in operational dimensions such as stack levels. As elaborated in the literature review, picking height has an established impact on order picking performance (Petersen et al. 2005). We thus integrate a moderator for the pick level. Table 3 summarizes our results for the interaction effects in Models (5) and (6) with separate interaction effects and in Model (7) for the full model. We present a visualization of the interaction terms in Figure 9.

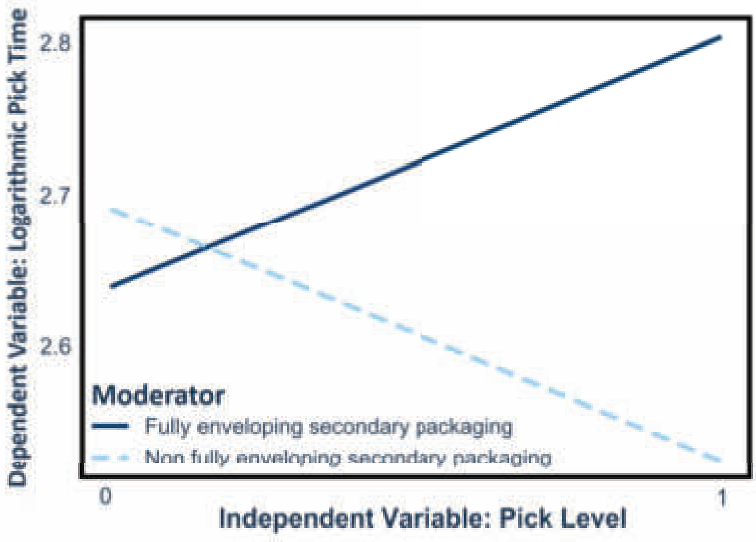

Figure 9: Interaction plot: Pick level and fully-enveloping secondary packaging.

Following the solid line in Figure 9, we find that for the case of fully enveloping secondary packaging, an increase in the pick level on the x-axis increases the pick time. This has an important implication for fully enveloping secondary packaging in the theoretical framework of operational-environmental performance.

While we initially defined fully enveloping secondary packaging as a Type-A trade-off located in Field 1 (positive effect for operational performance, negative effect for environmental performance), increases in the pick level can increase pick time and, therefore, negatively affect operational performance. Hence, we learn from the interaction term in Figure 9 that the Type-A trade-off for fully enveloping secondary packaging can lead to a transition into an undesirable situation when the picking level exceeds an empirically derived cross-over point.

The opposite applies to non-fully enveloping secondary packaging. When following the dashed line in Figure 9, we find that an increase in pick level (on the x-axis) decreases pick time for non-fully enveloping secondary packaging. That means that for non-fully enveloping secondary packaging pick time is reduced, when picking at chest level instead of picking close to the floor. While non-fully enveloping secondary packaging was initially defined as a Type-B trade-off, located in Field 3 (negative effect on operational performance, positive effect on environmental performance), increases in the pick level can decrease pick time and, therefore, positively affect operational performance. Hence, we learn from the interaction term in Figure 9 that the Type-B trade-off for non-fully enveloping secondary packaging can result in a transition to a win-win situation when the picking level exceeds an empirically derived cross-over point (pick level = 2, in our case). The non-filled circles in Figure 10 show the position of fully and non-fully enveloping secondary packaging when including pick level in the proposed theoretical framework.

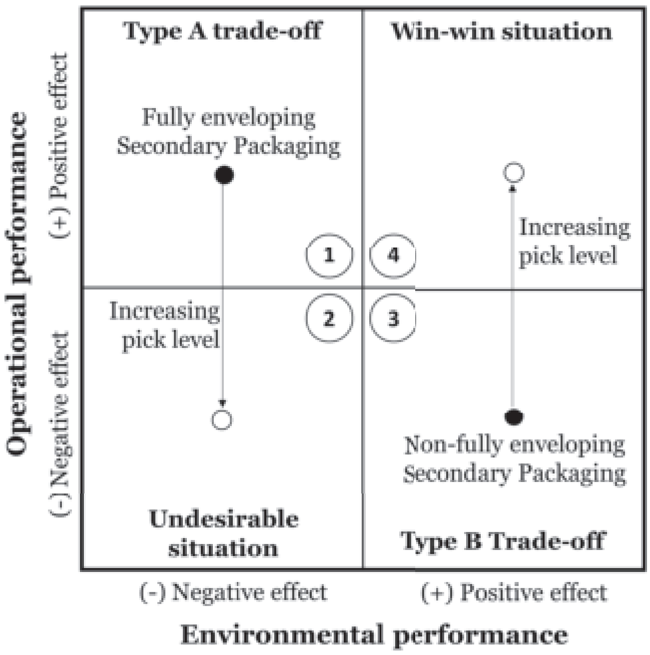

Figure 10: Trade-off positions of fully and non-fully enveloping secondary packaging.

We also check for the moderating effect of stack level, quantifying the height an order picker drops a stock-keeping unit after retrieving it from the storage location. Herein, we find a significant interaction effect with varying slopes of the solid and dashed lines in Figure 11. Given that we find no cross-over, the positions of fully and non-fully enveloping secondary packaging do not change in the theoretical framework for the interplay of operational and environmental performance when stacking levels vary. These findings are also good news for managers as they can directly affect pick height but hardly impact stacking height, which depends on product volumes and pick routes.

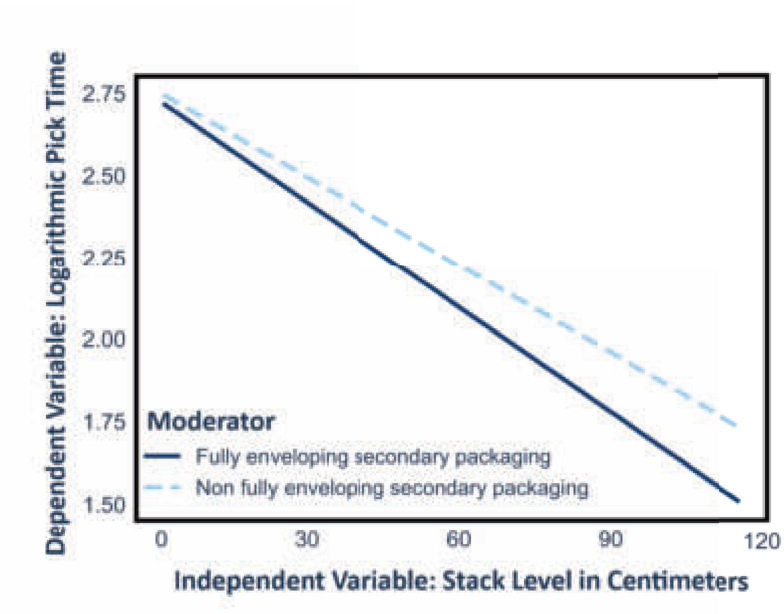

Figure 11: Interaction plot: Stack level and fully-enveloping secondary packaging.

For the one-piece secondary packaging characteristics, we identify a significant effect on operational performance. With a positive coefficient in Model (2), we can derive a negative effect on operational performance, which might result from systematic differences between one-piece secondary packaging and that consisting of several pieces. These systematic differences could result from omitted variables, potentially pointing to endogeneity issues. Endogeneity refers to the problematic scenario where an explanatory variable in a regression model is correlated with the error term, potentially leading to biased and inconsistent estimates (Lu et al. 2018). To reduce the risk of systematic differences, we add interaction terms in Models 3 to 7 and find that the main effect of one-piece secondary packaging changes when interacting with product volume (Model 4), stack level (Model 5), or pick level (Model 6). We conclude that the impact of one-piece secondary packaging on pick time can not be observed in isolation and highly depends on product characteristics (e.g., product weight and volume) and elements of warehouse design (e.g., pick and stack levels) moderating its impact. In addition, our data do not allow us to quantify the effect on environmental performance. Therefore, we assign one piece of secondary packaging to Fields 2 and 3 and several pieces of secondary packaging to Fields 1 and 4; we leave the exact assignment, which has an environmental dimension, to further research.

7 Conclusion

7.1 Implications for theory

Our study aimed to contribute to three major streams in the operations and logistics management literature: packaging logistics, order picking performance, and human workers in order picking systems. Furthermore, it contributes to the discussion on the interplay of economic and environmental performance. Our distinct contribution to packaging logistics is providing empirical evidence of the direct impact of secondary-level product packaging characteristics on operational warehouse performance metrics (Pålsson 2018, Sonck-Rautio et al. 2024). Our findings reveal that fully enveloping secondary packaging and carton box thickness notably reduce pick time. While acknowledging the endogeneity risks associated with the choice of carton box thickness, we mitigate these concerns by including a comprehensive set of control variables and interaction terms in our models. This approach helps to isolate the impact of carton box thickness on operational performance, reducing the potential bias from correlated unobserved factors.

Second, our study has important implications for the literature on order picking performance. The operations and logistics management literature suggests that optimizing tertiary packaging, which groups products into full unit loads for storage and shipping, is a central avenue for improving order picking performance (Hanson et al. 2018). While most brick-and-mortar retailers handle products at the secondary packaging level for store order fulfillment, this level remains largely under-explored (Freichel et al. 2020). Drawing on the different directions of the effects (positive and negative impacts on pick time), we can conclude that estimating the effect of secondary-level packaging characteristics requires a detailed decomposition rather than universal evaluation approaches.

Third, our study explores the impact of secondary packaging on operational and environmental performance in warehouses. We find that fully enveloping packaging improves operational efficiency but harms environmental performance (Type-A trade-off). In contrast, non-fully enveloping packaging is environmentally beneficial but less efficient operationally (Type-B trade-off). The effects of packaging on operational performance vary with factors like pick and stack levels, with changes in these factors potentially leading to undesirable or win-win scenarios. Carton thickness positively impacts operational performance but negatively affects the environment, representing another Type-A trade-off. Finally, when evaluating one-piece secondary packaging, we underline the need to consider product and warehouse characteristics.

7.2 Management implications

Retailers face intensified order-fulfillment challenges due to external pressures, such as increasing labor costs and workforce shortages. Consequently, warehouse managers actively seek ways to enhance performance and offset or neutralize these costs. Recent shifts toward inter-organizational sustainability initiatives add another layer of pressure. While a significant portion of food packaging is made of non-returnable materials, worsening consumers' material consumption footprint, the pursuit of corporate sustainability goals nevertheless has the potential to revolutionize product packaging norms.

We find that fully enveloping secondary packaging and carton thickness notably reduce pick time, potentially translating into substantial yearly savings given the level of European labor costs. Therefore, when corporate sustainability goals involve reducing packaging material (e.g., not fully enveloping secondary packaging or thinner carton boxes), we expect this will have negative implications for order picking performance and recommend that logistics managers incorporate this trade-off when formulating sustainability goals. Although we study the impact of product packaging characteristics on order picking performance in a retailer warehouse, we also advise taking a boundary-spanning supply chain perspective: (1) reducing material in tertiary packaging can increase material handling costs for suppliers, and (2) reducing material in secondary packaging may also impact store's handling costs when replenishing shelves.

We advise managers wishing to mitigate the adverse effects of product packaging characteristics by considering design variables, such as pick level, when assigning products to storage locations. This recommendation is grounded on the significant interaction effect we identify, with pick level moderating the impact on pick time of whether secondary packaging is a single piece. Here, we can conclude that placing one-piece secondary packing on low pick levels and packaging consisting of more than one piece on higher levels yields the best performance. In addition, we find that placing fully enveloping secondary packaging on low pick levels and non-fully enveloping secondary packing at a higher level also yields superior performance.

Finally, having spent time in the retailer's warehouse for field observations and picking items, we can confidently make several assumptions regarding the mechanisms behind these results. First, considering, for example, a tray of beer cans, whether the cans are enveloped in plastic makes a significant difference. Corporate sustainability initiatives may target reducing the plastic used in secondary packaging, but using alternative materials might contribute more to corporate sustainability and efficient operations. As a result, managers should consider how sustainable packaging choices might affect the speed and accuracy of order picking operations and their firm's overall material consumption.

Second, while product packaging characteristics might be directly aligned with retailers' private branding, branded products may require that procurement contracts and collaboration with suppliers be adapted to identify alternative materials. Finally, creating closed-loop supply chains for product packaging might be applicable to the majority of retailer's fruit and vegetable assortments, e.g., as applied for reusable secondary packaging in Europe with the Euro Pool System and the IFCO System.

7.3 Limitations and further research

Our study's limitations are mainly related to the applied usage of secondary data. While our work with a specific brick-and-mortar grocery retailer allowed us access to data on 4,957 dry food products, we were unable to broaden our analysis by including data from other retailers for an industry-wide perspective. We remain confident that our main results are robust because many retailers have a similar product assortment and, therefore, similar storage and picking structures. Our partner is a full-range European grocery retailer known for its broad product mix. In contrast to what might be expected for discount grocery retailers, product packaging variability in our study is not restricted to lower-priced items, potentially influencing the costs and quality of packaging materials. Further, we observe non-cooled perishable product categories familiar to many retailers, and our findings should thus be generalizable to similar contexts involving high variability in product packaging.

There are also limitations concerning the generalizability of our findings rooted in the order picking process. We study manual picker-to-parts order picking systems with a completely standardized process. Our findings should thus be generalizable to similar contexts, common in practice, in which order picking occurs on the ground level with the support of industrial trucks. However, some of our results could be dependent on the standardized process for subsequently visiting storage locations, limiting their applicability for situations where order picking tasks are not standardized and involve, for example, picking from heights higher than two meters by vertically moving industrial trucks or replenishing ground-level storage locations from the reserve in addition to ground-level retrieval.

We also examine an order picking system in which each picker handles one batch per route. Our insights should apply to comparable scenarios where products are retrieved from storage and sorted into groups of previously picked products. Nevertheless, some of our results may depend on sorting packaging characteristics into those for previously picked products. As a result, our findings may not be generalizable to pin-packing tasks in which one product is stacked into one bin. Additionally, these findings may not be accessible for replication studies where each picker handles more than one batch per route since this might involve additional search tasks (Batt and Gallino 2019).

Our work paves the way for future research on secondary-level product packaging characteristics. We take the first step in studying the role of product packaging characteristics on order picking performance in manual picker-to-parts order picking systems. Future work could test the extent to which our findings apply to hybrid or fully automated order picking tasks (Azadeh et al. 2019). Companies are investing in technologies to determine the exact volume, weight, and shape of product packaging before it enters their warehouses. This has two major benefits. Furthermore, future research could explore the performance benefits of standardized product packaging. Closed-loop supply chains, such as the IFCO System, can be expensive for retailers due to the need to wash and transport empty containers from the retailer back to the producer. However, these costs might be offset by performance improvements in the retailer's warehouses, potentially altering existing business models.

First, automated systems mostly rely on high data quality, including product master data, and exact volume, weight, and shape data are necessary for their efficient operation (Fragapane et al. 2021). Additionally, logistic service providers offering groupage or parcel services often employ product packaging data (e.g., on product volume, weight, and shape) in their freight pricing. As customers might intentionally or unintentionally declare data that deviates from the actual product packaging, primary packaging data are increasingly relevant. However, there is little empirical evidence on the business value of product packaging data quality across all levels.

Acknowledgements

The authors are grateful for the kind and productive exchange during the 2022 CSCMP European Research Seminar and the support by the editors and reviewers of the Journal of Business Logistics. Furthermore, we sincerely thank the team from Cambridge Proofreading for their high-quality support in proofreading and language editing. This research work was partly supported through the HumanTech Project, funded by the Italian Ministry of Universities and Research (MUR) for the 2023-2027 period as part of the initiative "Departments of Excellence" (L. 232/2016).

References

Azadeh, K., R. De Koster, and D. Roy (2019). Robotized and automated warehouse systems: Review and recent developments. *Transportation Science 53*(4), 917–945.

Bates, D. (2022). Linear mixed-effects models using 'eigen' and s4. version 1.1-33. *https://cran.r-project. org /web/packages/ lme4/lme4.pdf*.

Batt, R. J. and S. Gallino (2019). Finding a needle in a haystack: The effects of searching and learning on pick-worker performance. *Management Science 65*(6), 2624–2645.

Bentahar, O. and S. Benzidia (2018). Sustainable supply chain management: Trends and challenges. *Transportation Research Part E: Logistics and Transportation Review 119*, 202–204.

Bernerth, J. B. and H. Aguinis (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology 69*(1), 229–283.

Bliese, P. (2022). multilevel: Multilevel functions. version 2.7. *https://cran.r-project. org/web/packages / multilevel/ multilevel.pdf*.

Boysen, N., R. De Koster, and D. Füßler (2021). The forgotten sons: Warehousing systems for brick-and-mortar retail chains. *European Journal of Operational Research 288*(2), 361–381.

Brogaard, L. K., A. Damgaard, M. B. Jensen, M. Barlaz, and T. H. Christensen (2014). Evaluation of life cycle inventory data for recycling systems. *Resources, Conservation and Recycling 87*, 30–45.

Calzavara, M., C. H. Glock, E. H. Grosse, A. Persona, and F. Sgarbossa (2017). Analysis of economic and ergonomic performance measures of different rack layouts in an order picking warehouse. *Computers & Industrial Engineering 111*, 527–536.

Chabot, T., R. Lahyani, L. C. Coelho, and J. Renaud (2017). Order picking problems under weight, fragility and category constraints. *International Journal of Production Research 55*(21), 6361–6379.

Chan, F., H. K. Chan, and K. L. Choy (2006). A systematic approach to manufacturing packaging logistics. *The International Journal of Advanced Manufacturing Technology 29*(9-10), 1088–1101.

Chenavaz, R. Y., S. Dimitrov, and F. Figge (2021). When does eco-efficiency rebound or backfire? an analytical model. *European Journal of Operational Research 290*(2), 687–700.

Corbett, C. J. (2023). Om forum—the operations of well-being: An operational take on happiness, equity, and sustainability. *Manufacturing & Service Operations Management*.

Cuervo-Cazurra, A., U. Andersson, M. Y. Brannen, B. B. Nielsen, and A. Rebecca Reuber (2016). From the editors: Can i trust your findings? ruling out alternative explanations in international business research. *Journal of International Business Studies 47*(8), 881–897.

Davis-Sramek, B., A. Scott, and R. G. Richey (2023). Editorial: A case and framework for expanding the use of model–free evidence. *Journal of Business Logistics 1*, 1–7.

Davis-Sramek, B., R. W. Thomas, and B. S. Fugate (2018). Integrating behavioral decision theory and sustainable supply chain management: Prioritizing economic, environmental, and social dimensions in carrier selection. *Journal of Business Logistics 39*(2), 87–100.

De Koster, R., T. Le-Duc, and K. J. Roodbergen (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research 182*(2), 481–501.

De Vries, J., R. De Koster, and D. Stam (2016). Aligning order picking methods, incentive systems, and regulatory focus to increase performance. *Production and Operations Management 25*(8), 1363–1376.

Erbetta, F., C. Bruno, and C. Pirovano (2023). Corporate sustainability and performance: An efficiency perspective. *Business Strategy and the Environment 32*(6), 2649–2661.

Esfahbodi, A., Y. Zhang, Y. Liu, and D. Geng (2023). The fallacy of profitable green supply chains: The role of green information systems (gis) in attenuating the sustainability trade-offs. *International Journal of Production Economics 255*, 108703.

Figge, F. and T. Hahn (2012). Is green and profitable sustainable? assessing the trade-off between economic and environmental aspects. *International Journal of Production Economics 140*(1), 92–102.

Finnsgård, C. and C. Wänström (2013). Factors impacting manual picking on assembly lines: an experiment in the automotive industry. *International Journal of Production Research 51*(6), 1789–1798.

Fragapane, G., R. De Koster, F. Sgarbossa, and J. O. Strandhagen (2021). Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda: Literature review and research agenda. *European Journal of Operational Research 294*(2), 405–426.

Freichel, S., J. Wollenburg, and J. K. Wörtge (2020). The role of packaging in omni-channel fashion retail supply chains –how can packaging contribute to logistics efficiency? *Logistics Research 13*(1), 1–20.

Gattiker, T. F., C. R. Carter, X. Huang, and W. L. Tate (2014). Managerial commitment to sustainable supply chain management projects. *Journal of Business Logistics 35*(4), 318–337.

Golicic, S. L. and C. D. Smith (2013). A meta–analysis of environmentally sustainable supply chain management practices and firm performance. *Journal of Supply Chain Management 49*(2), 78–95.

Govindan, K., A. Rajeev, S. S. Padhi, and R. K. Pati (2020). Supply chain sustainability and performance of firms: A meta-analysis of the literature. *Transportation Research Part E: Logistics and Transportation Review 137*, 101923.

Grosse, E. H. and C. H. Glock (2015). The effect of worker learning on manual order picking processes. *International Journal of Production Economics 170*, 882–890.

Grosse, E. H., C. H. Glock, and W. P. Neumann (2017). Human factors in order picking: A content analysis of the literature. *International Journal of Production Research 55*(5), 1260–1276.

Gu, J., M. Goetschalckx, and L. F. McGinnis (2010). Research on warehouse design and performance evaluation: A comprehensive review. *European Journal of Operational Research 203*(3), 539–549.

Hahn, T., F. Figge, J. Pinkse, and L. Preuss (2010). Trade–offs in corporate sustainability: you can't have your cake and eat it. *Business Strategy and the Environment 19*(4), 217–229.

Hanson, R. and C. Finnsgård (2014). Impact of unit load size on in-plant materials supply efficiency. *International Journal of Production Economics 147*, 46–52.

Hanson, R., L. Medbo, M. Assaf, and P. Jukic (2018). Time efficiency and physical workload in manual picking from large containers. *International Journal of Production Research 56*(3), 1109–1117.

Hanson, R., L. Medbo, P. Jukic, and M. Assaf (2016). Manual picking from large containers - time efficiency and physical workload. *IFAC-PapersOnLine 49*(12), 1703–1708.

Jabbour, A. B. L. d. S., C. J. C. Jabbour, H. Latan, A. A. Teixeira, and J. H. C. de Oliveira (2014). Quality management, environmental management maturity, green supply chain practices and green performance of brazilian companies with iso 14001 certification: Direct and indirect effects. *Transportation Research Part E: Logistics and Transportation Review 67*, 39–51.

Kotzab, H., H. M. Munch, B. de Faultrier, and C. Teller (2011). Environmental retail supply chains: when global goliaths become environmental davids. *International Journal of Retail & Distribution Management 39*(9), 658–681.

Kumar, M. and V. S. Rodrigues (2020). Synergetic effect of lean and green on innovation: A resource-based perspective. *International Journal of Production Economics 219*, 469–479.

Lee, H.-Y. and C. C. Murray (2019). Robotics in order picking: evaluating warehouse layouts for pick, place, and transport vehicle routing systems. *International Journal of Production Research 57*(18), 5821–5841.

Loske, D. (2022). Empirical evidence on human learning and work characteristics in the transition to automated order picking. *Journal of Business Logistics 43*(3), 302–342.

Lu, G., X. Ding, D. X. Peng, and H. Hao–Chun Chuang (2018). Addressing endogeneity in operations management research: Recent developments, common problems, and directions for future research. *Journal of Operations Management 64*(1), 53–64.

Mahapatra, S. K., T. Schoenherr, and J. Jayaram (2021). An assessment of factors contributing to firms' carbon footprint reduction efforts. *International Journal of Production Economics 235*, 108073.

Mao, Z., S. Zhang, and X. Li (2017). Low carbon supply chain firm integration and firm performance in china. *Journal of Cleaner Production 153*, 354–361.

Masae, M., C. H. Glock, and E. H. Grosse (2020). Order picker routing in warehouses: A systematic literature review. *International Journal of Production Economics 224*, 107564.

Matusiak, M., R. De Koster, and J. Saarinen (2017). Utilizing individual picker skills to improve order batching in a warehouse. *European Journal of Operational Research 263*(3), 888–899.

McCoach, D. B. and A. C. Black (2012). Introduction to estimation issues in multilevel modeling. *New Directions for Institutional Research 2012*(154), 23–39.

Neumann, W. P. and J. Dul (2010). Human factors: Spanning the gap between om and hrm. *International Journal of Operations & Production Management 30*(9), 923–950.

Otto, S., M. Strenger, A. Maier-Nöth, and M. Schmid (2021). Food packaging and sustainability – consumer perception vs. correlated scientific facts: A review. *Journal of Cleaner Production 298*, 126733.

Pagell, M. and A. Shevchenko (2014). Why research in sustainable supply chain management should have no future. *Journal of Supply Chain Management 50*(1), 44–55.

Pålsson, H. (2018). *Packaging logistics: Understanding and managing the economic and environmental impacts of packaging in supply chains*. London and New York and New Delhi: Kogan Page.

Pan, J. C.-H., M.-H. Wu, and W.-L. Chang (2014). A travel time estimation model for a high-level picker-to-part system with class-based storage policies. *European Journal of Operational Research 237*(3), 1054–1066.

Petersen, C. G., C. Siu, and D. R. Heiser (2005). Improving order picking performance utilizing slotting and golden zone storage. *International Journal of Operations & Production Management 25*(10), 997–1012.

Schoenherr, T., S. B. Modi, S. Talluri, and G. T. M. Hult (2014). Antecedents and performance outcomes of strategic environmental sourcing: An investigation of resource–based process and contingency effects. *Journal of Business Logistics 35*(3), 172–190.

Seuring, S. and M. Müller (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production 16*(15), 1699–1710.

Sheffi, Y. (2018). Sustainability in practice. *Journal of Business Logistics 39*(3), 160–163.

Sonck-Rautio, K., T. Lahtinen, and N. Tynkkynen (2024). Consumer meaning -making of packaging functions for sustainable food packaging – insights from qualitative research in finland. *Current Research in Environmental Sustainability 7*, 100259.

Statista (2022). Statutory minimum hourly wages in countries of the european union.

Sun, J., D. J. Zhang, H. Hu, and J. A. van Mieghem (2022). Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science 68*(2), 846–865.

Tompkins, J., J. A. White, and Y. A. Bozer (2010). *Facilities planning* (4th edition ed.). Hoboken, NJ: Wiley.

UN (2024). The 17 goals. *online: https://sdgs.un.org/goals (2024/07/01)*.

Ünal, E. and V. K. Sinha (2023). Sustainability trade-offs in the circular economy: A maturity–based framework. *Business Strategy and the Environment 32*(7), 4662–4682.

Underwood, R. L. (2003). The communicative power of product packaging: Creating brand identity via lived and mediated experience. *Journal of Marketing Theory and Practice 11*(1), 62–76.

U.S. Bureau of Labor Statistics (2023). Compensation in private industry and state and local government. *online: https://www.bls.gov/charts/employment-cost-index/compensation-in-private-industry-and-state-and-local-government-12-month-percent-change.htm (2024/07/01)*.

U.S. Chamber of Commerce (2023). America works data center: Capturing the current state of the u.s. workforce. *online, https://www.uschamber.com/workforce/america-works-data-center (2024/07/01)*.

Vanheusden, S., T. van Gils, K. Ramaekers, T. Cornelissens, A. Caris, A. Caris, A. Caris, and A. Caris (2022). Practical factors in order picking planning: state-of-the-art classification and review. *International Journal of Production Research 61*(6), 2032–2056.

Walmart (2024). Sustainable packaging goals. *online: https://www.walmartsustainabilityhub.com/waste/sustainable-packaging/goals (2024/07/01).*

Wänström, C. and L. Medbo (2008). The impact of materials feeding design on assembly process performance. *Journal of Manufacturing Technology Management 20*(1), 30–51.

Zhu, Q. and J. Sarkis (2004). Relationships between operational practices and performance among early adopters of green supply chain management practices in chinese manufacturing enterprises. *Journal of Operations Management 22*(3), 265–289.

A.1 Correlation analysis

	1	2	3	4	5	6	7	8	9	10	11
Pick_Time											
One_Piece_Secondary_Packaging	0.03***										
Fully_Enveloping_Secondary_Packaging	0.00***	-0.26***									
Carton_Box_Thickness	-0.01***	-0.08***	0.06***								
Product_Weight	0.10***	0.36***	-0.08***	-0.02***							
Picker_Experience	-0.03***	-0.02***	-0.05***	0.00**	-0.06***						
Travel_Distance	0.26***	0.03***	-0.03***	0.00***	0.12***	0.03***					
Pick_Level	-0.02***	-0.18***	0.13***	-0.02***	-0.29***	0.02***	-0.04***				
Picks	0.24***	0.08***	-0.02***	-0.03***	0.12***	-0.03***	0.01***	-0.06***			
Stack_Level	-0.04***	-0.31***	0.14***	0.08***	-0.38***	-0.04***	-0.15***	0.11***	-0.07***		
Product_Volume	0.07***	-0.02***	0.14***	0.14***	0.29***	-0.03***	0.10***	-0.12***	0.03***	0.14***	
Primary_Packaging	0.03***	-0.11***	-0.01***	0.05***	0.10***	0.00*	0.00***	0.02***	-0.02***	0.10***	0.09***