Quantifying heterogeneity in human behavior: An empirical analysis of forklift operations through multilevel modeling

D. Loske¹,² · M. Klumpp¹,²,³

ABSTRACT

Human operators will remain to play an essential role in picker-to-parts order picking systems despite increasing digitalization and automation of warehouse processes. While manual order picking is a laborious and cost-intensive task in warehousing, it is extensively examined in the logistics and supply chain management literature. However, the operational and individual performance of forklift operators in warehouse picking operations has received little attention yet. We aspire to close this gap by drawing on sociotechnical systems theory and formulate a multilevel approach to evaluate heterogeneity in human behavior towards differences in picking performance. We use batch execution times as the dependent variable and source level of operation, target level of operation, filling level of the palette, the necessity to correct replenishment quantities, as well as travel distance as independent variables on the first level. For the second level, we utilize forklift operators to quantify whether heterogeneity in human behavior is impacting the performance of forklift operators. We find that 15.1% of the variance among batch execution times results from heterogeneity in human behavior. In a further simulation, we show that this method can be used to assess the performance of order pickers through a multi-dimensional parametric production frontier analysis. Our findings are highly relevant for logistics management when aspiring to forecast the necessary capacity of forklift operators in a warehouse or building bonus systems that are based on more than the existing two-dimensional measures such as process time per operation.

KEYWORDS: Human factor · behavioral operations management · forklift operators · multilevel modeling

1. INTRODUCTION

Intralogistics processes are characterized by ongoing discussions regarding further automation and digitalization [1–4], which include new robotics technologies [5], e.g., robotic sorting systems [6], autonomous mobile robots [7], or automated guided vehicles [8]. However, it is becoming increasingly obvious that the human factor is a highly relevant input factor for the performance of logistics systems, for example, in warehousing or order picking [9–11]. Despite this increasing relevance, analytical approaches and models, as well as empirical data on the performance impact of heterogeneity in human behavior in intralogistics operations, are very scarce. Three explanatory approaches might explain this existing gap: First, digital performance documentation has only recently become available to track individual intralogistics movements of forklifts with individual employees. Second, union and personal influence, as well as data protection regulations, might have prohibited a thorough investigation of personal performance inputs and developments. Third, management and leadership, as well as logistics research attention, might have been occupied with other topical fields such as technological development and large-scale logistics operations optimization.
from transport to location planning. Nevertheless, the identified research gap is highly relevant because great effort is implemented to attract suitable warehouse and intralogistics personnel, but detailed knowledge about which individual competencies are actually in high demand and highly efficient for intralogistics processes is lacking. It is an interesting objective for logistics research how the heterogeneity in human behavior can be measured for their contribution to overall efficiency and performance.

Therefore, we aspire to answer the following research questions to address this identified gap: (1) What are the relevant factors that significantly impact the batch execution times of forklift operators? (2) How and to what extent are differences among the batch execution times of forklift operators quantifiable and traceable to the heterogeneity in human behavior? The relevance of these questions becomes obvious when considering the role of human factors in manual picker-to-parts order picking systems [12, 13]. From a theoretical viewpoint, we draw on sociotechnical systems theory [14], a sub-field of general systems theory [15], and are especially interested in the behavioral aspects of human-technology collaboration [16].

This paper is structured as follows: The literature framework in Section 2 positions this study in the area of manual picker-to-parts order picking systems, highlights the theoretical background of this study, and presents existing research in behavioral operations management (BeOM) and human-centric analyses on forklift operation. Finally, parametric models that measure the efficiency of human operators in warehouses are presented. In Section 3, we describe the data and sample used in this study and introduce the multilevel model together with our strategy for validity and reliability. In the empirical results in Section 4, we discuss the impact of heterogeneity in human behavior and develop the final model in hierarchical steps. Based on the β-values of our final model, we formulate a multi-dimensional production frontier to evaluate operators’ performances in Section 5. Finally, managerial learning, limitations, future research direction, and a summary are presented in Section 6.

2. THEORETICAL FRAMEWORK

2.1. Forklift operations in manual picker-to-parts order picking systems

Warehouses play an essential role in most of the existing retail supply chains [14], and they impact the retailer’s business success to a large extent because they control product flows to meet customer demands [17]. Based on current figures on the global share of retail sales, e-retail sales accounted for 18% in 2020 [18]. Despite impressive growth rates in e-commerce, stationary retailing still accounts for the majority of global retail sales. However, as recent warehousing literature focuses on e-commerce, this proportion is not mirrored in recent scientific discussions [19]. Moreover, the underlying circumstances between warehousing in e-retailing and brick-and-mortar retailing are different. E-commerce warehouses mostly face low-volume-high-mix orders, as private households order few pieces from a large assortment provided by the online retailer [19]. An example is the average order demand at Amazon warehouses in Germany, which amounts to 1.6 pieces per order [20]. In contrast, the distribution centers of brick-and-mortar retailers process high-volume-low-mix orders. Thus, we position our paper in the area of warehousing for traditional brick-and-mortar retailing.

Because a detailed review of existing warehousing literature is beyond the scope of this paper, the reader is referred to [21–25] for systematic reviews with a general view on warehousing. [26–28] for reviews on performance measurement in warehousing, [29–31] for robotization and automation in warehousing, [19] for e-commerce warehousing and [32] for e-fulfillment in multi-channel distribution.

Main operations in brick-and-mortar retail logistics include receiving, transferring, and storing incoming items from suppliers for order picking and shipping to grocery stores [33]. Regarding the existing warehouse system, different order characteristics, including order size, assortment, workload variation, load stability, store-specific build-up, product expiry, and lead time mainly determine what is applicable for a given retail setting [20, 34]. Thus, suitable order picking systems may range from fully-automated case picking systems to fully manual picker-to-parts order picking systems. However, most of the existing warehouses in western Europe still adopt traditional picker-to-parts setups in their operations [35] due to the following facts: First, manual picker-to-parts order picking systems are highly flexible and scalable, as further order picking capacity can be generated by integrating additional order pickers and forklifts into the work system. Second, large assortments in retailing lead to high variability of products and special requirements for grasping and stacking the items (for example, fragility), which makes the picking task difficult to automate. In summary, our paper belongs to the research stream of manual picker-to-parts order picking systems where pickers travel with vehicle support.

Within the warehouses of traditional brick-and-mortar retailing, order picking is a necessary task, as incoming items from suppliers are usually received and stored in the form of large unit loads, for example, pallets. Customer orders originating from grocery stores are organized in fixed delivery cycles [36], and they comprise different items that mirror the assortments offered in the store while the volume is limited to the stores’ shelf capacity [37]. Meanwhile, a process that retrieves items from the large unit loads in response to a specific customer request is required, which is defined as order picking [38]. While picking operations are mostly done on ground level, the upper
2.2. Systems theory, sociotechnical systems and human-system interaction

Triggered by the introduction of computational technology in the mid-20th century, scholars and practitioners have been increasingly confronted with large-scale problems. The goal of integrating similarities within science, promoting communication across disciplines, and establishing a common theoretical basis for general scientific discussions and education have led to system thinking grounded in system science [15]. General systems theory enables the abstraction of large-scale problems by simplifying them on a common theoretical basis while simultaneously capturing the multi-dimensionality of the underlying problem [39]. The theoretical systems view has been applied to various circumstances where humans interact with other systems existing in reality, including socio-ecological systems interlinking human (social) and nature (ecological) [40] or socio-political systems [41]. As a first step, we position our research in systems theory.

In the 1950s, the transformation of working conditions led to hardly predictable work environments where interactions between humans and technology had to be considered [42]. The first set of principles for a sociotechnical systems design was proposed by Cherns in 1976 [43], which interlinked humans (social) and technology (machines). Since then, sociotechnical systems thinking has often been applied to human-machine work systems [14]. It has also been used to explain more modern phenomena such as human-computer systems [44] and has been extended to cyber–sociotechnical systems interlinking intelligent software (cyber) and humans (social) as well as machines (technical) [45]. In summary, sociotechnical systems theory is the theoretical basis that deals with the interaction of humans with any kind of engineered system [46]. However, an interaction such as a reciprocal action or influence is not trivial and may differ depending on several dimensions, such as workspace, work time, or the overall aim of the work system. The taxonomy applied by [16] systemizes these interactions as follows: human–system coexistence sharing work time and workspace; human–system cooperation sharing work time, workspace and aims; and human–system collaboration sharing work time, workspace, aims, and constantly in contact. In our case of forklift operation in manual picker-to-parts order picking systems, the human operator and machine share work time and workspace, and they both aim to fulfill retrieval operations to supply picking places at the ground level and fulfill storage operations from the ground level to the reserve level.

Additionally, they are constantly in contact because the operator sits in the forklift and utilizes the steering wheel to maneuver and control the machine parts to lift and lower forks. Hence, we are especially interested in the human-technology interaction of forklift operators.

2.3. Human–technology collaboration in forklift operation through the lens of human factors

As a scientific discipline, the human factor (synonymous with the term “ergonomics”) is concerned with the investigation of the interaction between humans and systems, or elements of a system, to optimize human well-being and overall system performance [47]. The theoretical basis is systems thinking and systems theory explained in the previous section, together with the claim that humans are part of operations and production systems and that they need to be integrated.

Calls for more human-centered research in these systems by [48], [49], [50], and more recently [51], point to the importance of human factors. The studies argued that deterministic models neglect the human factor by treating humans and technology, either implicitly or explicitly, as independent from each other. In recent reviews [9, 12, 13, 35], the human factor was divided into perceptual aspects (information processing, reading), mental aspects (learning, forgetting, behavior), physical aspects (workload, lifting, carrying), and psychosocial aspects (motivation, stress, feedback, monotony). However, the impact of these aspects on the forklift operator’s task is difficult to assess on a general level as mental and physical reactions to a perception depend on the respective tasks. Meanwhile, [46] argued that the application of the perception–cognition–motor–action cycle framework is relevant for all human-system interactions.

The cycle includes (1) gathering information through the human sensory system, (2) proceeding with the sensory input cognitively, (3) planning the action of the musculoskeletal system, and (4) responding with an action [52]. This action leads to a system output that is again perceived by the human operator. An existing schema in the human brain is then reinforced, extended, or newly built. This perception-action cycle is grounded on the circular cybernetic flow of cognitive information, a basic biological principle that links humans to their environment [53]. Table 1 systemizes the process steps of forklift operation and the relevant aspects of human factors for a detailed view of human–technology collaboration in forklift operation.
Research on human behavior and the human factor in forklift operation

Research on human behavior in operations systems aims to contribute to the field of BeOM, which has become an accepted sub-field of operations management [54, 55] and is relevant for supporting managers in decision making, designing and optimizing employees’ working conditions, as well as improving processes for customers [56]. While BeOM starts at the micro-level to facilitate a better understanding of the heterogeneity in human behavior, the research approaches associated with this stream analyze decisions, skills, and capabilities of individuals or small groups [57, 58]. This may include a cognitive perspective, for example, individual risk aversion in the context of the newsvendor problem [59], individual cognitive reflection in the same context [60], individual perception [61], social preferences [62], or trust in technology [63–65]. The physical perspective is rarely addressed by BeOM, but it is extensively addressed by the more engineering-driven human factor stream. Although forklift operators are less studied than order pickers, there is still a considerable body of research dealing with them in the context of human-factor analysis. [66] examined cold-related symptoms among forklift drivers operating in a cool warehouse in Thailand. Further approaches that deal with the physical factor explore neck pain [67], neck and backloads [68], energy consumption [69], the impact of whole-body vibration [70], or low back pain [71]. Additionally, a well-examined sub-process of forklift operation is the driving process itself, which is the reason many studies investigate collision warning systems [72, 73], safety systems [74], and sensor-based assistant systems [75].

### Table 1: Forklift sub-processes through the lens of human factors

<table>
<thead>
<tr>
<th>Operator tasks</th>
<th>Perceptual</th>
<th>Mental</th>
<th>Physical</th>
<th>Psychosocial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setup</strong></td>
<td>Perceiving the setup of the operation, e.g., batch details on a touch display.</td>
<td>Receive and understand the setup, e.g., batch details on a touch display.</td>
<td>Movements when setting up the workstation, e.g., bending or neck flexion extension during a departure check.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perceive warehouse layout, including the visual record and orientation in aisles, shelves, and places.</td>
<td>Understand and remember a route when starting at a source and traveling to a target location.</td>
<td>Steer the forklift through the warehouse, including, e.g., arm movement while sitting or standing in front of a steering wheel.</td>
<td></td>
</tr>
<tr>
<td><strong>Travel</strong></td>
<td>Perceive arrangement of places with visual record and orientation with the place logic of the warehouse</td>
<td>Search and identify locations and items, e.g., load unit in target level.</td>
<td>Movements during the search process, including neck flexion extension when, e.g., searching the environment.</td>
<td>Aspects regarding motivation, stress, workload, boredom, and interaction with co-workers within all operator tasks, including the setup, travel, search, as well as store/retrieve tasks.</td>
</tr>
<tr>
<td><strong>Search</strong></td>
<td>Perceive height of racks with visual record and orientation when operating within several rack levels.</td>
<td>Decide on how to handle sub-processes, e.g., how to position the forklift and how high to lift forks during a retrieval operation.</td>
<td>Movements during the search process, including neck flexion extension when, e.g., bending for optimal view on forks.</td>
<td></td>
</tr>
<tr>
<td><strong>Retrieve/store</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Examples in brackets. Initially introduced by [35] for order picking.
well examined. [76, 77] examined fatigue associated with the sub-process of driving. An approach that contributes to this direction is exploring the situation awareness of forklift operators for preventing collisions [78]. Hence, we identify a research gap regarding the interplay of perceptual, mental, and physical aspects when humans operate forklifts within a human–technology setting. This may include the fact that some individuals are more efficient in combining task sets than others. As humans are an essential part of these work systems, they significantly impact the outcomes of warehouse systems, including performance and quality. In the next section, we will focus on methodologies to operationalise, evaluate and benchmark human-operator performance in warehousing.

2.5. Parametric models for performance measurement of human operators in warehousing

When multilevel modeling is applied to warehousing, the basic idea is to measure the performance of order picks and decompose large datasets that contain order batches. However, it is often difficult to determine how and to what extent human capabilities and skills cause performance differences. In one approach, [10] applied multilevel modeling to order picking logs and proved that almost 10.3% of the total time variance in batch execution time could be traced to individual differences. Another approach was presented by [79] using pick duration as the dependent variable. They analyzed the impact of the independent variables through a survival model, more precisely a parametric accelerated-failure-time. More parametric models have been applied to measure fatigue and learn order pickers [80, 81], using Wright’s, Sandford-B, or DeJong’s learning curve. Most of these approaches use cumulated picks and investigate the development of the time per pick while concluding that minimizing the pick time is equal to a learning effect where a human operator improves picking skills.

3. DATA AND METHODOLOGY

3.1. Warehouse and dataset description

We analyzed the performance of forklift operators employed in the distribution center of a large German grocery retail chain. Within this distribution center, refrigerated and non-refrigerated foods are stored, picked, packed, and delivered to supermarkets. This study focuses on one warehouse storing perishable and non-perishable non-cooled items. In our sample warehouse, orders are picked through a manual picker-to-parts order picking system, and storage locations used for picking are replenished by forklift operators. The warehouse has two cross-aisles, 28 aisles on the ground level, one depot storing empty roll cages, and drop-off points for the pickers at every aisle. Since the racking system comprises several gripping levels (minimum, ground level; maximum, 1.05 m) and five reserve levels (level 1, 2.10 m; level 2, 4.20 m; level 3, 6.30 m; level 4, 8.40 m; level 5, 10.5 m), the storage of pallets in the higher levels is particularly a challenge. The maximum height for the delivered pallets (including pallet timber) is 1.05 m (CCG1) or 1.95 m (CCG2). Altogether, storing or retrieving a 1.95-m CCG2 pallet in a storage place with a height of 1.95 m on the highest rack level (10.5 m) is a challenging task. This is why we expect individual heterogeneity and behavioral factor to be relevant for the batch execution time. Many warehouse management systems (WMS) store replenishment and order picking logs, which quantify operations on a detailed level. The batch execution times for individual forklift operators may depend on warehouse-specific circumstances, for example, a varying amount of order pickers operating in the same warehouse, which depends on work time and workgroup decisions. However, as human capabilities, skills, and heterogeneity in human behavior have been identified as relevant for intralogistics [12, 80], behavioral factors may also influence batch execution times.

We included these behavioral factors implicitly through the available data by considering and evaluating the past performance of the forklift operators. One batch is equal to a storage or retrieval operation of one single full pallet. The warehouse management data obtained for this study on batch execution times for replenishment operations of forklift operators included: (1) batch ID, (2) forklift operator ID, (3) starting date of the batch, (4) ending date of the batch, (5) starting time of batch, (6) ending time of batch, (7) distance traveled by a forklift as a difference between the source and target locations, (8) height of the pallet, (9) level of the source location, (10) level of the target location, (11) degree of filling for each pallet as a percentage of the CCG1 or CCG2 type, (12) necessity of correcting the amount on a picking location as errors occur in upstream processes (as a dichotomy variable, 0 = no and 1 = yes).
Table 2.: Descriptive statistics of the observed forklift operators

<table>
<thead>
<tr>
<th>Operator ID</th>
<th>Observations</th>
<th>Mean execution time</th>
<th>Standard deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,198</td>
<td>155.63</td>
<td>79.74</td>
<td>0.51</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>161.39</td>
<td>67.14</td>
<td>0.42</td>
</tr>
<tr>
<td>3</td>
<td>735</td>
<td>114.65</td>
<td>67.35</td>
<td>0.59</td>
</tr>
<tr>
<td>4</td>
<td>908</td>
<td>190.81</td>
<td>91.19</td>
<td>0.48</td>
</tr>
<tr>
<td>5</td>
<td>626</td>
<td>199.17</td>
<td>91.16</td>
<td>0.46</td>
</tr>
<tr>
<td>6</td>
<td>469</td>
<td>176.10</td>
<td>99.66</td>
<td>0.57</td>
</tr>
<tr>
<td>7</td>
<td>920</td>
<td>112.53</td>
<td>70.55</td>
<td>0.63</td>
</tr>
<tr>
<td>8</td>
<td>673</td>
<td>171.85</td>
<td>102.11</td>
<td>0.59</td>
</tr>
<tr>
<td>9</td>
<td>961</td>
<td>118.21</td>
<td>67.84</td>
<td>0.57</td>
</tr>
<tr>
<td>10</td>
<td>446</td>
<td>177.35</td>
<td>68.66</td>
<td>0.39</td>
</tr>
<tr>
<td>11</td>
<td>1,022</td>
<td>179.73</td>
<td>87.82</td>
<td>0.49</td>
</tr>
<tr>
<td>12</td>
<td>1,357</td>
<td>147.11</td>
<td>75.45</td>
<td>0.51</td>
</tr>
<tr>
<td>13</td>
<td>932</td>
<td>107.78</td>
<td>43.05</td>
<td>0.40</td>
</tr>
<tr>
<td>14</td>
<td>981</td>
<td>207.23</td>
<td>89.98</td>
<td>0.43</td>
</tr>
<tr>
<td>15</td>
<td>1,261</td>
<td>193.05</td>
<td>111.66</td>
<td>0.58</td>
</tr>
<tr>
<td>16</td>
<td>1,154</td>
<td>111.07</td>
<td>48.71</td>
<td>0.44</td>
</tr>
<tr>
<td>17</td>
<td>467</td>
<td>250.46</td>
<td>122.42</td>
<td>0.49</td>
</tr>
<tr>
<td>18</td>
<td>423</td>
<td>195.05</td>
<td>104.30</td>
<td>0.54</td>
</tr>
<tr>
<td>19</td>
<td>945</td>
<td>157.13</td>
<td>66.67</td>
<td>0.42</td>
</tr>
<tr>
<td>20</td>
<td>537</td>
<td>194.59</td>
<td>98.72</td>
<td>0.51</td>
</tr>
<tr>
<td>21</td>
<td>1,167</td>
<td>183.64</td>
<td>74.00</td>
<td>0.40</td>
</tr>
<tr>
<td>22</td>
<td>778</td>
<td>153.20</td>
<td>98.58</td>
<td>0.64</td>
</tr>
<tr>
<td>23</td>
<td>789</td>
<td>153.40</td>
<td>87.72</td>
<td>0.57</td>
</tr>
<tr>
<td>24</td>
<td>431</td>
<td>222.00</td>
<td>102.23</td>
<td>0.46</td>
</tr>
<tr>
<td>25</td>
<td>633</td>
<td>139.48</td>
<td>91.98</td>
<td>0.66</td>
</tr>
<tr>
<td>26</td>
<td>635</td>
<td>210.72</td>
<td>96.94</td>
<td>0.46</td>
</tr>
<tr>
<td>27</td>
<td>706</td>
<td>99.92</td>
<td>38.57</td>
<td>0.39</td>
</tr>
<tr>
<td>28</td>
<td>1,129</td>
<td>204.17</td>
<td>110.12</td>
<td>0.54</td>
</tr>
<tr>
<td>29</td>
<td>1,275</td>
<td>165.98</td>
<td>102.15</td>
<td>0.62</td>
</tr>
<tr>
<td>30</td>
<td>525</td>
<td>185.00</td>
<td>96.88</td>
<td>0.52</td>
</tr>
<tr>
<td>31</td>
<td>980</td>
<td>169.92</td>
<td>75.97</td>
<td>0.45</td>
</tr>
<tr>
<td>32</td>
<td>707</td>
<td>125.07</td>
<td>72.03</td>
<td>0.58</td>
</tr>
<tr>
<td>33</td>
<td>709</td>
<td>193.94</td>
<td>85.01</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Quantifying heterogeneity in human behavior: An empirical analysis of forklift operations through multilevel modeling

The whole dataset contained 27,985 batches (equal to full loads on pallets) for replenishment operations performed by 45 forklift operators during a 1.5-month period. All operations were performed between 7 a.m. and 6 p.m., and all the operators used identical technology. As we use real-life data, some datasets were polluted by personal breaks during batches. Therefore, we filtered out all batches lasting longer than 15 mins and all batches without operator ID. Additionally, we excluded zero values and checked for unreasonable traveling speed within the warehouse. After the data cleaning, the dataset containing 26,979 out of 27,985 batches and 33 out of 45 forklift operators were used for further analysis. Table 2 summarizes the descriptive statistics of the operators.

### 3.2. Multilevel modeling

Linear regression models are used to explain the statistical relationship between a dependent variable \( y_i \) and an independent variable \( x_i \). They can be extended to multiple linear regression with more than one independent variable. Therefore, one possible approach is to explain the batch execution time of a forklift operator for batch \( i \) as the dependent variable and the travel distance and filling level of a pallet, both on the individual batch level \( I \), as independent variables. The standard ordinary least squares (OLS) regression, which includes the error term \( e_i \), is given by Eq. (1):

\[
y_i = \beta_0 + \beta_1 x_i + e_i
\]  

(1)

Multilevel modeling depicts a series of nested OLS regression analyses in which the coefficients at level 1 become the dependent variable at another level [82]. In the first level of our model, we tried to explain the batch execution time with the five independent variables used to quantify heterogeneity in human behavior. Table 3 summarizes the dependent and independent variables for the first level of our model to operationalize behavioral factors.

In our multilevel model, we represent the sub-tasks of the individual operator by linear regression model betas. Let \( \beta_{ij} \) be one sub-task within \( i \in I \), where \( i \) is the number of the examined task, and \( J \) is the total set of tasks described in Table 3. Additionally, \( j \in J \) defines the forklift operator with an individual operator \( j \) and a set of examined operators \( I \). After clarifying the basic idea of our approach based on multiple linear regression, we added a second level. Multilevel modeling is used to additionally distinguish between groups [83–85], which are represented by forklift operators in our case. Herein, we added group variables that enabled us to differentiate between groups and to evaluate whether or not this differentiation is beneficial for the model. Furthermore, we concretized the dependent variable \( y_{ij} \) as the batch execution time \( t \) for a set of possible batches \( r \in R \) from the set of order \( O \) that we also used for the independent variables. The hierarchical two-level model was notated as a group-dependent regression, with \( j \) reflecting the group level. We used the notation of [10] formulating a multilevel approach for order pickers to forecast the picking time in the first level and grouping according to operator ID:

\[
t_{r,w} = \beta_{0,w} + \sum_{i \in I} \beta_{i,w} x_{i,r} + e_{w,r}
\]  

(2)

\[
\beta_{0,w} = \gamma_{1} + u_{0,w}
\]  

(3)

### Table 3.: Operationalization of human behavior and variables for level 1

<table>
<thead>
<tr>
<th>Behavioral factor</th>
<th>Role in model</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>The performance of the forklift operator shall be high when the processing time for storage and replenishment operations is minimized</td>
<td>Dependent</td>
<td>Batch execution time</td>
</tr>
<tr>
<td>The heterogeneity in the speed of work execution when performing the sub-task of retrieving pallets from the racking system with different levels</td>
<td>Independent (( \beta_1 ))</td>
<td>Source level of operation</td>
</tr>
<tr>
<td>The heterogeneity in the speed of work execution when performing the sub-task of storing pallets on the grasping level for the order picker</td>
<td>Independent (( \beta_2 ))</td>
<td>Target level of operation</td>
</tr>
<tr>
<td>The heterogeneity heterogeneity in the speed of work execution when performing the sub-task of handling different heights of pallets where the difficulty increases with high filling levels (less space for maneuveringmanoeuvring)</td>
<td>Independent (( \beta_3 ))</td>
<td>Filling level of a pallet</td>
</tr>
<tr>
<td>The heterogeneity in the speed of work execution when performing the sub-task dealing with errors and correct them through his mobile computer connected to the WMS</td>
<td>Independent (( \beta_4 ))</td>
<td>Necessity to correct</td>
</tr>
<tr>
<td>The heterogeneity in the speed of work execution when performing the sub-task of driving a forklift fast and saving within the assigned warehouse area</td>
<td>Independent (( \beta_5 ))</td>
<td>Travel distance</td>
</tr>
</tbody>
</table>
After calculating the null model as a basis for further evaluation in Section 4.1, we used the model fits in Eqs. (4)–(6) to assess whether or not the integration of additional dependent variables is beneficial for the total $R^2$. This approach is similar to hierarchical regression analysis and enables the evaluation of integrated level one predictors.

4. EMPIRICAL RESULTS

4.1. Differences between operators

In the first step, we calculated a variance components model without predictors $x_{i \text{op}}$, which is well-known as a null model in multilevel modeling [82]. Herein, one fixed effect of the grand mean, in our case, the mean of the batch execution times, was estimated for all batch IDs and operator IDs. The null model is given as follows:

$$
\tau_{r,w} = \beta_{0,w} + \varepsilon_{w,r}
$$

$$
\beta_{0,w} = y_i + u_{i,w}
$$

Based on this, we computed the intraclass correlation coefficient (ICC) to evaluate between-group variations [87–89]. In our case, this measure can explain whether or not variations in batch execution times occur between forklift operators and if it is suitable to apply multilevel modeling. For the mathematical formulation, the ICC is given by [82]:

$$
ICC = \frac{\sigma^2_{\tau_{op}}} {\sigma^2_{\tau_{op}} + \sigma^2_{\varepsilon_{ij}}}
$$

where

- $\sigma^2_{\tau_{op}}$: variance of batch execution time in the second level (operator ID)
- $\sigma^2_{\varepsilon_{ij}}$: variance of batch execution time in the first level (batch ID)

The ICC for our case is 0.1510, indicating that 15.1% of the variance among the batch execution time originates from the forklift operators on the second level of our multilevel model, and 84.9% can be traced back to the batches themselves. [10] conducted the same analysis for order pickers, and their model showed that 10.3% of the variance in batch execution time was due to the individual differences of order pickers. The results are presented in Table 4, and they confirm that our results are significant at the alpha level of 0.05. As the Wald-Z value of 4.03 is highly significant ($p < .001$), we can reject the null hypothesis that there are no significant differences in batch execution time among the examined forklift operators.
Quantifying heterogeneity in human behavior: An empirical analysis of forklift operations through multilevel modeling

4.2. A two-level model with fixed predictors and random intercepts

In a second analysis stage, we built on the null model and added travel distance per batch as the first independent variable on the first level. Additionally, we added the operator IDs in the second level. For the initial bi-variate linear regression model with execution time as the dependent variable, we find a high statistical linear relationship with travel distance as the independent variable. This is also illustrated in Figure 1 with linear regression lines per forklift operator.

The individual and different slopes per operator indicate the high heterogeneity in human behavior observable for the examined forklift operators. Furthermore, we integrated this independent variable, together with the other independent variables introduced in Table 2, into our multilevel approach. For the integration of travel distance, we formulated the level 1 in Equ. (10) and the level 2 in Equ. (11, 12) as follows:

\[
t_{r,w} = \beta_{0,w} + \beta_{1,w} \text{travel.distance}_{i,r} + \epsilon_{w,r} \quad (10)
\]
\[
\beta_{0,w} = \gamma_1 + u_{i,w} \quad (11)
\]
\[
\beta_{1,w} = \gamma_{10} \quad (12)
\]

Table 4.: Summary of null-model for intraclass correlation coefficient test

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Wald-Z</th>
<th>p</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_{e_{ij}})</td>
<td>7293.55</td>
<td>62.84</td>
<td>116.07</td>
<td>.000</td>
<td>7171.43</td>
<td>7417.75</td>
</tr>
<tr>
<td>(\sigma_{u_{ij}})</td>
<td>1297.91</td>
<td>322.144</td>
<td>4.03</td>
<td>&lt;.001</td>
<td>797.95</td>
<td>2111.13</td>
</tr>
</tbody>
</table>

Note: The lower and upper bounds are for an alpha level of 0.05.
Independent variables. Additionally, the integration of operator IDs as a second level increases the null model up to 41%. This again confirms that heterogeneity in human behavior is a critical factor. In total, our final model increased the predicting power up to 86% compared to the initial null model. Considering the model betas, the results for the source level of operation (β₁) are not significant and must be excluded from our analysis. Therefore, our final model comprises target level of operation (β₂), filling level of the pallet (β₃), the necessity to correct (β₄), and travel distance (β₅).

In summary, our findings show that heterogeneity in human behavior plays a major role when forecasting batch execution times with these variables. Table 6 summarizes all model betas for a multilevel model with fixed predictors and random intercepts.

### Table 5: Summary of a multilevel model with fixed predictors and random intercepts

<table>
<thead>
<tr>
<th>Model</th>
<th>( R^2 )</th>
<th>( R_1^2 )</th>
<th>( R_2^2 )</th>
<th>Parameter</th>
<th>Estimate</th>
<th>AIC</th>
<th>Wald-Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null-model</td>
<td></td>
<td></td>
<td></td>
<td>( \sigma_{\epsilon_i} )</td>
<td>7293.65</td>
<td>316703</td>
<td>116.07</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma_{\mu_i} )</td>
<td>1297.91</td>
<td>4.03</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>.84</td>
<td>.92</td>
<td>.38</td>
<td>( \sigma_{\epsilon_i} )</td>
<td>609.42</td>
<td>249802</td>
<td>116.07</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma_{\mu_i} )</td>
<td>800.29</td>
<td>4.06</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>.86</td>
<td>.94</td>
<td>.41</td>
<td>( \sigma_{\epsilon_i} )</td>
<td>402.69</td>
<td>238638</td>
<td>116.06</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma_{\mu_i} )</td>
<td>769.74</td>
<td>4.06</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>.86</td>
<td>.94</td>
<td>.41</td>
<td>( \sigma_{\epsilon_i} )</td>
<td>402.59</td>
<td>238634</td>
<td>116.07</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma_{\mu_i} )</td>
<td>769.73</td>
<td>4.06</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Model 5</td>
<td>.86</td>
<td>.94</td>
<td>.41</td>
<td>( \sigma_{\epsilon_i} )</td>
<td>400.78</td>
<td>238513</td>
<td>116.07</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma_{\mu_i} )</td>
<td>764.68</td>
<td>4.06</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Model 6</td>
<td>.86</td>
<td>.94</td>
<td>.41</td>
<td>( \sigma_{\epsilon_i} )</td>
<td>400.63</td>
<td>238505</td>
<td>116.07</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( \sigma_{\mu_i} )</td>
<td>764.25</td>
<td>4.06</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** AIC is used as an estimator to predict error rates in the model.

Compared to the initial null-model, our model 2 (null-model + travel time as dependent variable) showed an improved Akaike information criterion (AIC), indicating a reduction of the error rate in our model. We added the additional independent variables stepwise until we reached the final model described in Eqs. (2) and (3) with model 3 (model 2 + the filling level of the pallet), model 4 (model 3 + source level of operation), model 5 (model 4 + target level of operation) and model 6 (model 5 + necessity to correct). Table 5 summarizes the results.

The results indicate that the level 1 prediction power of the final model (model 6) increases by 94% when integrating the variables described above. This finding on predicting the batch execution time is not surprising, and it still proves that it can be forecasted with the given independent variables. Additionally, the integration of operator IDs as a second level increases the null model up to 41%. This again confirms that heterogeneity in human behavior is a critical factor. In total, our final model increased the predicting power up to 86% compared to the initial null model. Considering the model betas, the results for the source level of operation (β₁) are not significant and must be excluded from our analysis. Therefore, our final model comprises target level of operation (β₂), filling level of the pallet (β₃), the necessity to correct (β₄), and travel distance (β₅). In summary, our findings show that heterogeneity in human behavior plays a major role when forecasting batch execution times with these variables. Table 6 summarizes all model betas for a multilevel model with fixed predictors and random intercepts.

### Table 6: Model betas for a multilevel model with fixed predictors and random intercepts

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \beta )</th>
<th>Standard error</th>
<th>df</th>
<th>p</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>50.93</td>
<td>4.835</td>
<td>33</td>
<td>.000</td>
<td>41.10</td>
<td>60.76</td>
</tr>
<tr>
<td>Source level of operation</td>
<td>-.01</td>
<td>.0110</td>
<td>26949</td>
<td>.484</td>
<td>.029</td>
<td>.014</td>
</tr>
<tr>
<td>Target level of operation</td>
<td>1.32</td>
<td>.155</td>
<td>26948</td>
<td>.000</td>
<td>1.016</td>
<td>1.62</td>
</tr>
<tr>
<td>Filling level of the pallet</td>
<td>119.80</td>
<td>1.031</td>
<td>26951</td>
<td>.000</td>
<td>117.79</td>
<td>121.83</td>
</tr>
<tr>
<td>Necessity to correct</td>
<td>1.22</td>
<td>.383</td>
<td>26947</td>
<td>.001</td>
<td>.47</td>
<td>1.97</td>
</tr>
<tr>
<td>Travel distance</td>
<td>.48</td>
<td>.002</td>
<td>26949</td>
<td>.000</td>
<td>.48</td>
<td>.49</td>
</tr>
</tbody>
</table>

**Note:** Degrees of freedom (df) is the number of values in the final model that are free to vary.
5. SIMULATION OF BATCH EXECUTION TIME FOR FORKLIFT OPERATORS

After formulating the multilevel approach and applying it to the real data of 33 forklift operators, we are now able to formulate a multiple regression model with the estimates calculated in the previous section. At this point, it is important to note that the source level of operation ($\beta_1$) is excluded in the simulation approach. From a managerial perspective, there are two possible tracks for simulation. First, by integrating level 1 and level 2 formulations in Eqs. (25) and (26), it is possible to forecast the batch execution times for forklift operators and the total working time for all batches. This may be beneficial when logistics managers seek to find the optimal capacity of forklifts and forklift operators for their warehouse. Second, after being aware that individual heterogeneity in human behavior has an impact on the performance of forklift operators, we adopt the level 1 regression model to forecast the batches. Hence, we can formulate a multi-dimensional and parametric production function that is beneficial to evaluate a forklift driver’s performance compared to the group performance level. The formulation of our simulation approach is as follows:

$$t_{r,w} = 50.93 + 1.32 x_{2,r} + 119.8 x_{3,r} + 1.44 x_{4,r} + 0.48 x_{5,r} + \varepsilon_{w,r}$$  \hspace{1cm} (13)

$$\beta_{0,w} = y_1 + u_{1,w}$$  \hspace{1cm} (14)

$$\beta_{1,w} = y_{10}$$  \hspace{1cm} (15)

*Fig. 2.* Comparison of simulation results and real batch execution times for forklift operators
After running the simulation, we found significant differences among the individual forklift operators, which is as expected from our previous findings and can be used for cross-validation purposes. As bonus schemes in intralogistics mostly rely on process time, for example, the average time per operation, a multi-dimensional parametric approach may show a promising way to integrate further performance-relevant factors. Additionally, our approach integrates time for correcting errors that result from upstream processes and are, therefore, not caused by forklift operators. However, these situations are often ignored when benchmarking blue-collar workers in instore logistics.

6. CONCLUSION

6.1. Management learnings
This paper has outlined an innovative method and empirical evaluation approach regarding the analysis of efficiency contributions from human operators with forklift operators in retail warehousing operations. A relevant explanation share of 15.1% regarding the individual batch operating time data was identified, warranting further research and also management attention towards the analysis of individual heterogeneity. Our model is transferable to all warehouses where forklift operations are performed by humans. This may include picker-to-parts or parts-to-picker order picking systems. The main requirement for transferring our formulated model to practice is the availability of historical data, e.g., documented by a WMS. Additionally, our method requires data on the aggregation level of operations. Because most modern warehouses utilize digitalized or intelligent picking systems, e.g., pick-by-voice or pick-by-vision, we expect that the data used in this paper is also available to a majority of warehouse managers. To further optimize warehouse processes, our findings may help managers to rethink batch assignment methods in human-centered work systems and reorganize work system design strategies.

First, taking the aspect of batch assignment methods, most companies use simple rules, such as the first free method. We expect a high application potential for batch assignment respecting the individual heterogeneity of the human workforce. After formulating the model, operationalizing the constructs, and inserting the data into SPSS or R, the software output provides model betas per operator. Through a comparative analysis of all operators, we are able to cluster humans according to their performance, e.g., operators that are better in driving the forklift than others. At the same time, we can cluster all batches in the WMS according to, e.g., distance. Bringing both together, we may utilize the model betas from our multilevel approach to implement human-centered batch assignment for forklift operations. This would also allow improving training processes for new and unknown employees: Let us assume that operator A has an experience level of 10,000 cumulative operations after a time period of three months and is seen as an experienced employee. Our model and the historic WMS data can now help to estimate the model betas in the very first week quantifying a low level of experience. These model betas can then be assigned to new and unknown operator B. In this exemplary case, the human-centered batch assignment may be suitable to prevent overload and resulting stress when incorporating new employees.

Second, data-driven multilevel analyses might be beneficial to develop human-centered work system design strategies. The presented parametric multilevel models have the potential to shed light on operations with human involvement at a very detailed level. Finding that 15.1% of the total variance in time is related to differences between human operators might, e.g., lead to incentive-based sociotechnical work systems taking behavioral factors more into account. It underlines the importance of integrating ergonomic aspects for work system design. Additionally, our method allows the evaluation of operators’ performance on sub-tasks, e.g., driving the forklift. This may be helpful to derive whether and to which extent assistive technologies may be necessary for certain tasks, sub-tasks, operators, or batches. For an extensive review of assistive devices in manual material handling, the reader is referred to [90]. In summary, quantifications of individual heterogeneity in operations systems might trigger managers to rethink work system design strategies.

6.2. Limitations and further research
The approach presented in this paper has specific limitations. First, although the obtained data basis of 26,979 batches remaining after the data cleaning is extensive, the quantifications are limited to the examined warehouse. Therefore, our model and the approach would benefit from further applications in alternative warehouse scenarios. Additionally, we investigate humans operating in a warehouse storing perishable and non-perishable non-cooled items. It may, therefore, be interesting to research the impact of temperature on human performance and heterogeneity in cold supply chains, as well as the performance in night shifts compared to those in day shifts.

From a methodological viewpoint, further research avenues can be directed at multilevel models for panel data. This could enable the quantification of the learning effects and the development of human skills at a certain time. Furthermore, non-parametric methodologies need to be tested to build production frontiers without setting regression weights a priori. Data envelopment analysis and free disposable hull methodologies are suitable as they apply a model-endogenous weighting of input and output factors. The importance of research analyses and concepts regarding the human-technology interaction parts of operations and logistics systems will further
increase due to future digitalization and automation developments.

From a content-oriented viewpoint, logistics managers may be interested in the aspect of how and to which extent technological improvement through digitalization and automation can accelerate operations. Further research utilizing the same historic WMS data as in our approach may, therefore, be directed in the area of parametric survival models. The event-history analysis is concerned with duration times where an event is defined as the transition from the origin to the destination state [91]. This time frame is quantified as a survival object and integrated as the dependent variable in the survival model. Thus, we propose to set the process time of interest as the survival object and the dependent variable within an accelerated failure time model. Based on this, independent variables can then be defined, allowing to quantify their accelerating or decelerating impact.

6.3. Summary and Outlook

The contribution of this paper is threefold: (1) it elaborates, justifies, and applies a parametric multilevel approach as a novel method for evaluating forklift operators, (2) it decomposes batch execution time variations in two levels to reveal the impact of heterogeneity in human behavior, (3) it elaborates and applies a simulation approach based on a parametric multi-dimensional production frontier to evaluate operators’ performance with empirical data. Therefore, this paper provides the first step and is a cornerstone for future analyses regarding the efficiency impact of the human factor. This is connected to ongoing research regarding the human factor in logistics operations [19, 30, 46, 81, 92–96]. The common ground of this research stream is the empirical evidence pointing at a highly relevant efficiency and effectiveness impact of human factors towards logistics operations, as also shown empirically in this contribution. This warrants further research and analysis of the relevant impact of logistics research.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES


Human Factors 46/1(1):50–80. doi:10.1518/hfes.46.1.50.30392


