Designing a compatibility evaluation framework for integration of 3PL warehouse clients in warehouse systems: an action design research study in a Danish 3PL company

Frederik H. Nielsen [ID],[✉], Anton S. Hermann [ID] and Niels Henrik Mortensen [ID]

Department of Civil and Mechanical Engineering, Technical University of Denmark, Denmark

✉ fhol@dtu.dk

ABSTRACT: This study explores the design of a compatibility evaluation framework for integrating 3PL warehouse clients into semi-automated warehouse setups. Using Action Design Research (ADR), an artifact was developed that combines data-driven decision-making (DDD) and multi-criteria decision analysis. The framework, implemented in Microsoft Power BI, enables the evaluation of client compatibility based on configurable criteria and relevant metrics. It was co-created with stakeholders and tested using data from 33 warehouse clients, demonstrating its practical value in identifying operational fit while facilitating data-driven discussions. The study highlights the potential of structured decision frameworks in environments with limited data, offering generalizable insights for 3PL warehouses and similar contexts.

KEYWORDS: Evaluation, Collaborative design, Action Design Research, Warehouse Systems, Organizational processes

1. Introduction

Warehouses are no longer just storage facilities; they have evolved into critical elements in company supply chains. They provide additional services like product display building, labelling, and cross-docking to meet diverse client needs. These advancements bring challenges in designing and operating warehouses that balance operation speed, flexibility, and the potential integration of automated systems (Gu et al., 2007, 2005; Baker and Canessa, 2009). This is especially a concern within the third-party logistics (3PL) sector, where clients outsource their logistics to 3PL suppliers, who in turn might have several clients in the same warehouse. As the 3PL industry grows, so do client requirements, and with it, market competition (Khakdaman et al., 2023). Consequently, 3PL clients are prone to be less loyal towards specific providers, with a tendency to value service offerings and low prices (Munsberg et al., 2022; Strøm et al., 2023a). To stay competitive while also profiting, warehouse 3PL providers need to ensure that client implementation is done correctly, with both warehouse design and operations fitting the client (Gu et al., 2005, 2007).

In this study, we focus on the design and implementation of a warehouse client data model containing an evaluation framework within a Danish 3PL company. At the case company, a decision was made to move clients from a smaller warehouse which contained mainly fashion and E-commerce clients. These types of customers are notably demanding, as many retailers are now integrated into both physical and online shops, often referred to as omni-channel retailing (Kembro et al., 2018). With the clients relocating, the case company requested a re-evaluation of their warehouse setup to account for the differences at the new, larger facility with automation options. Therefore, a series of client evaluations were requested to

avoid situations where a potential efficiency gain through automation could risk getting undermined by the increased complexity and inefficiencies arising from misaligned integration.

This paper addresses these challenges by proposing and validating a compatibility evaluation framework for assessing warehouse clients to determine the best-fitting setup. Compatibility is in this context understood as the similarity to the 'ideal' client for a specific warehouse solution (Akarte and Ravi, 2007). Through the use of Action Design Research (ADR) methodology, iterative design cycles were conducted, in which a warehouse client data model was introduced to the industry partner. This data model includes a configurable multi-criteria analysis with an evaluation algorithm, and is considered the main *artefact*. Through the artefact, warehouse clients are ranked based on compatibility with a standardized semi-automated warehouse setup, serving as a foundation for decision-making. To better explore conditional evaluations, the underlying algorithm was made configurable to quickly examine different scenarios., which received praise from both practitioners and end-users. The paper concludes with a reflection on the lessons learned during the design process, presented as concrete design principles, along with a broad evaluation of the artefact.

1.1. Literature review

Warehousing serves as a hub for storing, handling, and distributing goods, often categorized as stockkeeping units (SKUs) (Gu et al., 2007). Core processes in a warehouse include the *Inbound* stage, where goods are received, *Storage*, where inventory is managed, and the *Outbound* stage, where goods are picked, packed, and shipped to the end customer (Strøm et al., 2023b). Beyond these standard operations, warehouses frequently perform a range of client-specific services i.e. cross-docking and order consolidation. This diversity in services contributes to significant operational variability across warehouses, as each facility must adapt its processes to meet the unique requirements of its clients (Munsberg et al., 2022).

This client variance contributes to the integration of new clients being inherently complex (Munsberg et al., 2022). The requirements of each client — ranging from the nature of their goods to their operational timelines — introduce challenges in aligning their needs with the warehouse's existing infrastructure (Baruffaldi et al., 2020). Moreover, the lack of standardized processes for onboarding new clients further complicates integration, leading to inefficiencies that can undermine the overall performance of the facility (Strøm et al., 2023a). Successful integration is essential not only for maintaining operational efficiency but also for fostering long-term client relationships in the highly competitive 3PL market.

A critical part of the integration is making sure that the most ideal warehouse solution is chosen for the client. Certain warehouse systems, such as semi-automated configurations or specific storage solutions, are better suited for specific types of clients. Misalignment between a client's needs and the warehouse setup can lead to increased operational complexity, inefficiencies, and ultimately higher costs (Gu et al., 2005, 2007), putting pressure on the initial solution-decision.

Customer-solution matching approaches have previously been explored through decision-making frameworks, evaluating *compatibility* between products or services and customer or production needs (Dahlberg, 1983; Akarte and Ravi, 2007; Xiao et al., 2022; Xiu and Chen, 2012). In the 3PL industry, compatibility models have primarily used to analytically compare providers to identify the most suitable match (Xiu and Chen, 2012; Jung, 2017). Similarly, Akarte and Ravi (2007) developed a framework to align casting products, processes, and producers with an *ideal foundry*, representing optimal capabilities. Such models consistently define compatibility as the alignment between evaluation criteria and desired attributes.

A common methodology employed in compatibility evaluations is a multi-criteria analysis using variable weights, which systematically assesses and ranks alternatives based on predefined metrics. This type of approach has been widely applied to evaluate operational complexity and optimize decisionmaking under uncertainty (Thurston, 1990; Akarte and Ravi, 2007; Cross, 2000). In the context of 3PL, weighted criteria frameworks have traditionally focused on evaluating providers from the customer perspective, such as 3PL providers ability to meet sustainability KPIs or deliver tailored logistics solutions (Jothimani and Sarmah, 2014; Jung, 2017). However, the reverse perspective—using client data to evaluate compatibility with warehouse setups—remains notably absent in the literature, as such marking a research gap.

Additionally, despite some advancements in warehouse research, much of the existing literature has been characterized as empirical and descriptive, with limited application of theoretical frameworks in practice (Prockl et al., 2012). Furthermore, the disconnect between academic research and warehouse

practitioners persists, hindering the development of actionable insights that could enhance industry practices (Gu et al., 2005).

This paper aims to address the identified gaps by proposing a compatibility evaluation framework tailored for 3PL warehouses. By leveraging weighted criteria analysis and operational data, the framework seeks to make the integration process more data-driven and transparent. In doing so, this research contributes not only to improving operational efficiency in 3PL settings but also to strengthening the collaboration between academia and practitioners in warehouse operations.

2. Designing an evaluation framework for 3PL warehouse setup client integration

This section details the design of an evaluation framework focusing on rating 3PL warehouse clients from a warehousing perspective, based on the evaluation principles from multi-criteria analysis research.

2.1. Design setting and background

As a means to design the evaluation framework, the Action Design Research (ADR) research method was utilized. The reasoning for the choice of ADR was the strong focus on generating prescriptive design knowledge through the creation and evaluation of IT artefacts within an organization (Thiess and Müller, 2020; Sein et al., 2011). In the context of this research paper, the ADR method was chosen for its compatibility with the project setting, as two of the authoring researchers were employed as industrial PhD students within a large Danish 3PL provider. This company will from hereon be referred to as the "Industry Partner" or "IP". Due to this setting, practitioners and end-users were always within reach in the shape of trusted colleagues, simplifying data gathering, co-creation, and feedback rounds.

The Action Design Research (ADR) method emphasizes the collaborative involvement of academics and industry experts to address real-world problems. Sein et al. (2011) introduces a structured four-stage process, each guided by specific principles: (1) Problem Formulation, (2) Building, Intervention, and Evaluation, (3) Reflection and Learning, and (4) Formalization of Learning.

The proposed ADR methodology is fundamentally iterative, focusing on cycling between the first three stages based on lessons learned from each of them. Supporting this method framework, Sein et al. (2011) details a generic schema for Organization-Dominant BIE (*building* of the IT artefact, *intervention* in the organization, and *evaluation*). In this schema, the "ADR Team" is described to consist of researcher(s) and organizational practitioners, working together to iteratively develop versions of an IT artefact, in which end-users provide feedback during this process, to assist in finalizing the artefact. The major iteration steps made in this project can be seen in Figure 1 inspired by this schema. With a strong focus on co-creation, researchers, practitioners, and end-users all add unique contributions to the design of the IT artefact. Sein et al. (2011) summarize these contributions as: Researchers contributing with design principles, practitioners contributing to the specific ensemble being design, and finally end-users adding utility for the users.

The Industry Partner's warehousing strategy is tied to following modern market trends, with a focus on innovation and competitiveness. To better meet client's unique demands, the IP's warehouses are equipped with specific storage setups/systems, including automated/semi-automated setups, setups for hazardous goods, "fast pick" areas, etc.. In the context of this study, two of the authors had been asked to assist the IP in an internal moving operation, which entailed transferring 33 warehouse clients from an older and smaller warehouse to a newer and modern warehouse. During this process, the IP desired to evaluate which setup in the new warehouse would be best suited for the clients being transferred. Problems arise when the nature of the client is unknown, as warehouse systems often are designed towards specific segments of goods (Gu et al., 2007). Furthermore, if operational deficiency has been detected, it has proven difficult for the IP to assess, which setup would be better suited for the client's goods. Nevertheless, as with most modern warehouses, a Warehouse Management System (WMS) is utilized to track, manage, and analyze most operations happening within the walls of the warehouse (Gu et al., 2007; Munsberg et al., 2022). For the older warehouse, a more simple WMS system had been used (referred to as WMS1 from now on), whereas the newer warehouse runs on a more advanced WMS system (WMS2). The historic WMS1 data provided an opportunity to investigate how general logistics client data could support improved decision-making for integrating clients into specific warehouse setups, which is introduced as part of the artefact design.

2.2. The design

Following the move-case, the IP formulated a problem formulation used in the first stage of the ADR method: "There is currently no implemented method or tool to evaluate the compatibility of warehouse setups and warehouse clients using historic warehouse data". With this formulation in mind and having access to historic data from WMS1, a data analysis effort was conducted, in which data from the 33 clients were analyzed. The data sample included 12 months of data for most clients, which consisted of inbound order data, outbound order data, as well as a list of all Stock-Keeping Units (SKUs) registered with that client. From these three databases, outbound order data was selected as the foundation for the analysis for three reasons:

1) From personal experience, order line data is great for determining client seasonality (what orders, and how many, go out at what time of the year, with how much quantity per order?)
2) The size and processing of each order can help determine what operational flow is best suited for the client (B2B or B2C operational considerations)
3) Outbounding (or Pick and Pack) has been documented as the most labour intensive, and as such a large cost-driving element of warehouse operations (de Koster et al., 2007)

Inspired by Geoff McGrath's emphasis on maximizing insight from minimal data (McGrath, 2022), the framework's design has been developed based on principles similar to Akarte and Ravi (2007) to evaluate compatibility. The goal was to identify relevant statistical criteria for determining the 'ideal' setup. The outbound order line database was therefore chosen as the basis for analysis due to the richness and familiarity of information it offers. The artifact development, documented in the following sections, involved multiple micro-iterations aligned with stages 1, 2, and 3 of the ADR method (Sein et al., 2011), with only major steps summarized in Figure 1.
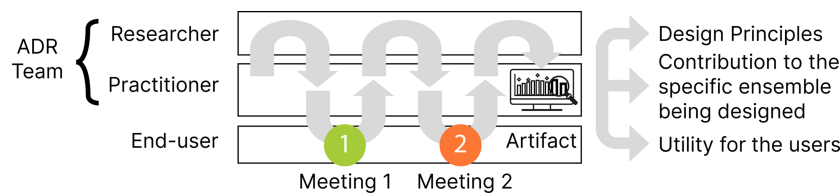

Figure 1. ADR Process and Iterations based on the Organization-Dominant BIE schema (Sein et al., 2011)

2.2.1. Initial data insights

From the first data insights, the second stage of the ADR method began with the creation of a data model in Microsoft Power BI. By using a data model, it is possible to combine, summarize, and compare the statistics of the clients included in the analysis. Furthermore, by using a company-wide utilized software such as Power BI, shareability within the organization was made easier. The initial inspection of the order data highlighted basic order line information was identified, which included **Order IDs, Order Date, SKU IDs** and **SKU Order Quantity**. From experience working with these data types, a variety of statistical insights could be made, including:

- Order size insights, including average and median SKU order quantity per order and order line (AQO & MQO)
- Seasonality insights such as overviews of which months had the most/least amount of orders, order lines, and/or order quantities (OS)
- SKU statistics such as nr. of distinct SKUs pr. client and average distinct SKUs pr. order (NDS & ASO)
- SKU volumetric data (VD) as a way to determine size requirements

These variables can later be seen summarized in Table 1. From an operational perspective, these statistics can help identify the typical order size, and order frequency, as well as how active their SKU usage is, e.g., what is the relationship between the number of distinct SKUs and the amount of orders. In setups in which SKUs are assigned storage locations allowing item picks, having large numbers of unique SKUs

and similar amounts of orders, would not be optimal. From a feedback round with a warehouse manager, the seasonality aspects were discussed, in which it came to attention, that certain automated warehouse setups will become inefficient if the system is exposed to too many orders at the same time. As such, if a client shares the same seasonality trend with multiple other clients in the same setup, operational efficiency could be affected negatively.

2.2.2. Meeting 1: Building on the analysis

The preliminary insights were shared continuously throughout the design process, but two central meetings with practitioners and data experts served as the main points of iteration. During the first meeting, the stakeholders provided constructive feedback on potential uses for the data and suggested additional statistical metrics for evaluating the clients. From the experts, who were familiar with the data structure and tags used in WMS1, it was possible to extract which orders had been handled as *cross-docking* orders (CD), which orders are consolidated (CSD), as well as identify some helpful product categories and outbound package parameters (PC & OPC). Clients having a large proportion of cross-docking orders do not benefit from setups that have a slow inbound process (Gu et al., 2007), and consolidation creates complexity, as some warehouse setups do not have inherent support for consolidation operations. Finally, using the outbound package data, one could determine the size of the carton the package was sent in, which indicates if it was most likely a B2C type order (e.g. plastic bag) or B2B order (e.g. full pallet).

Table 1. Overview of variables included in evaluation framework

Variables	Influence	Relative Weight	Characteristics
Cross Docking % (CD)	↓	High	B2B Operations
Volumetric Data % (VD)	↑	High	Container Turnover Rates
Average Shipped QTY pr. Order(line) (AQO)	↓	Medium	Container Turnover Rates
Median Shipped QTY pr. Order(line) (MQO)	↓	Medium	Container Turnover Rates
Nr. of distinct SKUs (NDS)	↓	Low	Storage Capacity
Average distinct SKUs pr. Order (ASO)	↓	Low	Storage Usage
Order Seasonality (OS)	↕	Low	Operation Intensity
Inbound Package Criteria (PC)	↕	High	Miscellaneous
Outbound Package Criteria (OPC)	↕	High	Miscellaneous
Nr. of distinct carton types (DCT)	↓	Low	Packing Complexity
Consolidation % (CSD)	↓	High	Packing Complexity

Table 1 shows a summary of the final variables considered for the client evaluation. Note that the naming of these variables and the number of variables have been simplified for this paper. In the case of IP's move project, the scope was to evaluate the client's compatibility with a semi-automated setup in the new warehouse. From multiple stakeholder meetings, it became apparent, that the semi-automated setup was best utilized for operations consisting of small order sizes typically connected to the B2C segment. The setup also had a dependency on specific inbound and outbound procedures that rely on SKU volumetric and characteristics. As such, Table 1's **Influence** column reflects how each statistical variable would influence the evaluation, e.g., an "↑" indicating that, the higher the number, the more positive the fit. For variables with "↕", the influence is more complex, for example, Order Seasonality depending on if the seasonality peaks are shared with other clients. The column **Characteristics** gives a simplified description of what warehouse aspect the variable resembles.

The column **Relative Weight** highlights how important each variable is deemed to be when evaluating a client's compatibility in the semi-automated setup. To assign the appropriate weights for each variable, the "Weighted Objectives Method" (WOM) was utilized (Cross, 2000). By comparing each variable with each other in a *WOM Matrix*, e.g. VarA and VarB, a score can be assigned of either; 1, if the first variable is more important than the other, 0, if it is the opposite, or 0.5 if they are deemed equally important. To assign the weights, subjective evaluations were made together with warehouse practitioners, in which

each variable was evaluated and ranked according to domain knowledge. The WOM matrix and final scores were adjusted over multiple iterations, with an example of the matrix shown in Table 2.

Table 2. Example of Weighted Objectives Method (WOM) matrix inspired by Cross (2000)

WOM	CD	VD	AQO	MQO	NDS	ASO	OS	PC	OPC	DCT	CSD	Score
CD		0.5	1	1	1	1	1	0.5	0.5	1	0.5	8
VD	0.5		1	1	1	1	1	0.5	0.5	1	0.5	8
AQO	0	0		0.5	1	0.5	0.5	0	0	1	0	3.5
MQO	0	0	0.5		1	0.5	0.5	0	0	1	0	3.5
NDS	0	0	0	0		0.5	0.5	0	0	0.5	0	1.5
ASO	0	0	0	0	0.5		0.5	0	0	0.5	0	1.5
OS	0	0	0.5	0.5	0.5	0.5		0	0	0.5	0	3.5
PC	0.5	0.5	1	1	1	1	1		0.5	0.5	0.5	7.5
OPC	0.5	0.5	1	1	1	1	1	0.5		0.5	0.5	7.5
DCT	0	0	0	0	0.5	0.5	0.5	0.5	0.5		0	2.5
CSD	0.5	0.5	1	1	1	1	1	0.5	0.5	1		8

Following the principles presented by Cross (2000), the WOM scores should be seen as a hierarchy of variables, in which the final *weights* would be found by ranking the *impact* of each variable relative to each other (e.g. by distributing the variables on a scale from 1-10). Having identified relevant criteria, and assigned appropriate weights, an *multi-criteria* evaluation algorithm was designed as an attempt to return a compatibility score.

The algorithm is based on the concept of comparability, and as such, some of them were translated to a normalized scale. For a few of the categories, a decision was made to make a relative comparison, which can be done to quickly compare relevant clients. It should be noted, that relative comparisons are not the most ideal practice, as they can easily be distorted by outliers within the client group (Thurston, 1990). Following the principles from Cross (2000) and the multi-attribute function by Thurston (1990), each criterion was multiplied by the decided weight. Finally, the sum of the clients' scores was divided by the maximum potential score to give a result between 0 % and 100 %.

In addition, a design choice was made to highlight good and bad scores with red and green. This gives an indication of which criteria would have to be improved for the client to increase their overall compatibility score. This score would indicate how closely the compared clients align with an 'ideal' client in the context of the warehouse setup. These final rankings and criteria would then be displayed in a table, such as the example shown in Table 3.

Table 3. Simplified Evaluation Table Example

Clients	Combined Evaluation	CD	VD	AQO	MQO	NDS	ASO	DCT	CSD
Client A	85%	0%	20%	2.02	1	4100	0.4	4	0%
Client B	77%	2%	26%	9.8	3	25200	0.25	11	18%
Client C	68%	10%	44%	2.15	1	10616	0.06	29	20%
Client D	53%	0%	0%	11.6	5	330	0.14	6	25%

Following the comparative algorithm, the artefact and resulting rankings of each client were showcased and discussed with practitioners and end-users. The feedback was generally positive as the evaluation tool was the first operations-data-based analytical tool introduced in the move-project, that could both indicate compatibility, as well as highlight areas of concern. One critical type of feedback that was received, was leading questions such as: "What if we solved the consolidation issues, how would that change the evaluation"? or "What if we only look at non-cross docking orders, would those SKUs be a great fit?". Essentially, to better answer hypothetical scenarios where different implementation strategies could be employed, a final iteration of the evaluation artefact was created.

2.2.3. Meeting 2: The inclusion of evaluation configuration

In the second meeting, a desire for dynamic client evaluation was proposed. The main wish was to make the evaluation tool meet end-users' desire to easily adapt the algorithm to specific *what-if* scenarios. Design-wise, that meant that functionality should be added that enabled the users to configure the implementation case they desired. Therefore, a new artefact iteration was created in which each criterion could be added or removed with ease. As an example, it had been discussed that every client would eventually have to have 100% volumetric data if the semi-automatic solution was selected. Going with this assumption, the **Volumetric Data % (VD)** variable should not be included in the evaluation of the client, and would therefore be removed from the configuration. Including the functionality of the Power BI data model filters, end-users would be able to filter the data to better suit the scenarios of interest, by creating multiple different analysis configurations. The addition of customizability marked the finalization of the artifact.

2.2.4. Intervention: Presentation and integration of the artefact

The final version of the artefact was presented to both practitioners and end-users within the case company. The feedback was mostly positive, with the inclusion of customizable evaluation criteria receiving very strong endorsement, as it enabled the end-users to quickly compare different implementation scenarios. When presented with the top-ranked clients in different configurations, some practitioners expressed that the ranking reflected their subjective understanding of the client's compatibility within the chosen setup. Other client rankings were perceived as unexpected, in which the different evaluation criteria were inspected to help explain the ranking. When shown criteria weighing down a client in focus, it was often met with a confirming understanding of the statistical variable, and with comments on how the client could or should be handled to create a better setup fit.

The artefact's main intervention was meant to affect the client integration process, as it was deemed the most important business process for the framework. As an intervention, the artefact was received positively, with stakeholders mentioning that the artefact enabled a better way to have data-based discussions and that although the evaluation score isn't perfect, it was better than any data-based tool previously utilized. Additionally, as part of the client integration process, the artefact was deemed useful as a supporting evaluation of the qualitative experiences used by practitioners when evaluating warehouse clients.

In terms of shareability, the artefact was designed in Microsoft Power BI following the IP's IT governance restrictions, which enabled both the data model and interactive dashboard to be shared within the organization. Overall, the artefact was deemed to be a useful contribution by the included end-users, and a desire was expressed to utilize the artefact in the continued moving project, as it was viewed as a vast improvement in data utilization. The result of the tool's utility in specific client evaluations is out of scope for this paper but could serve as inspiration for future research.

3. Reflection and learning

The framework's reliance on simple, scalable metrics ensures broad applicability beyond the IP's WMS, offering a structured approach to addressing complexity drivers in warehousing, aligning with methodologies utilized by other researchers (Kembro et al., 2018; Akarte and Ravi, 2007). However, the use of relative comparison is not ideal. We, the researchers consider the relative comparisons a temporary necessary evil but understand the potential negative impacts it might have, which the customizability of the artefact might increase. By making certain configurations of the analysis practitioners and end-users might enhance the potential negative effect of the relative comparisons, creating an unwarranted negative bias. Such bad practices have not yet been observed. Still, the risk underlines the need for future research to validate these criteria across diverse settings and establish empirical links to operational costs.

Additional future work should explore the framework's validity across more and different types of clients, potentially integrating more relevant criteria. The criteria incorporated in the artefact were based on insights from domain experts as well as the limitations of the data source i.e. type of clients and database quality. Hence, by including more domain experts and other data sources, other needs might become apparent, improving the level of insight without challenging the need for applicability. Another noteworthy element is to include more data per client to increase the robustness of the seasonality analyses in alignment with the desire to have at least two full seasonality cycles of data (IBM, 2024; Dataiku Knowledge Base, 2024).

More generally, the design was done by having two industry PhD researchers working with the IP. While this approach offers valuable insights, it also introduces the potential risk of bias due to close proximity to the industrial context. To negate this, a strong focus was put on following the ADR methodology, constantly looking at the general scientific contributions, which can be deducted from the final learnings and design principles.

In regards to the design process of the artefact, the ADR methodology was instrumental in emphasizing stakeholder collaboration and iterative artefact development. However, the challenge of documenting micro-iterations highlights an area for methodological improvement, aligning with findings from (Thiess and Müller, 2018). Addressing this limitation in future ADR applications could provide a more comprehensive representation of the iterative design process while further enhancing the alignment between academic and practical insights.

4. Formalization of learning

This study highlights the challenge of client analyses based on limited or incomplete data generated for operation, a challenge common across many industries (Sebastian-Coleman, 2022). In this study, we encountered such constraints while designing the evaluation framework for client compatibility. While the specifics of our data and use case were tied to the IP's systems, and on a larger scale, the 3PL warehousing industry, the core approach to addressing client analysis challenges has broader applicability.

As a result, a widely recognized methodology was adopted. The application of multi-criteria analysis has been extensively documented as effective across various disciplines, making it a logical choice for the development of this framework for evaluating 3PL clients. This approach incorporates the customization of analysis and leverages low scores as a basis for identifying areas of improvement, aligning with similar frameworks (Thurston, 1990; Akarte and Ravi, 2007; Cross, 2000). However, no prior studies were identified that specifically focused on client-centric evaluation within 3PL warehousing. For this reason, we retain that this paper provides a valuable contribution both to the theory covered, and to the industrial practice.

4.1. Design principles

The following design principles, derived from our iterative design process, aim to provide generalizable insights that can guide practitioners and researchers in similar contexts where systematic, data-driven decisions are required under constraints.

4.1.1. Design principle 1: Simplify for scalability and comparability

Aligning with popular design principles such as KISS (Interaction Design Foundation, 2024), this principle addresses that scalability and comparability are achieved by leveraging straightforward and comparable statistics that can be executed on basic databases, such as an *order line database*. We, the researchers, recommend others to limit this type of analysis to widely applicable databases, especially if one wishes to screen potential new clients in the same manner.

4.1.2. Design principle 2: Ensure transparency of evaluation foundations

Transparency in evaluation frameworks is crucial for increasing meaningful discussions with stakeholders. Clearly defining variables and their weightings in the interface of the artefact facilitates comprehension and encourages constructive feedback from practitioners and end-users. This aligns with findings by Thiess and Müller (2020), who argue for theoretical over complex models in ADR studies, based on the likelyhood of theoretical models being closer with stakeholders' mental models.

4.1.3. Design principle 3: Prioritize customizability in evaluation tools

Customizability enhances the usability of evaluation tools by allowing stakeholders to adapt analyses to specific scenarios. This principle addresses the need for dynamic configurations to explore what-if scenarios and analysis outcomes aligning with insight from Akarte and Ravi (2007). The experience from this project was that assumptions changed into actionable strategies for value-creation, enabling stakeholders to identify and plan against any potential warehouse issues.

Together, these principles emphasize the importance of simplicity, transparency, and adaptability in designing compatibility-evaluation frameworks.

5. Conclusion

In this paper, the design of a warehouse client system evaluation framework for use within a Warehouse 3PL provider was covered. Warehousing no longer covers just simple storage operations, but is a critical part of a supply chain, with the adoption of complex client flows. To meet demands from clients in competitive markets, more focus is being put on the utilization of data and technology. An internal decision to move warehouse clients, motivated the development of a client evaluation framework within a Danish 3PL provider.

Based on the utilization of the Action Design Research (ADR) methodology, a final artifact was designed, featuring a configurable table of key statistics derived from a simple order-line database, enabling real-time exploration of analysis scenarios. Based in multi-attribute analysis principles (Thurston, 1990; Akarte and Ravi, 2007), the design was praised for its customizability, which facilitated data-driven discussions and reduced reliance on intuition. Stakeholders considered it a significant improvement in decision-making processes.

Additionally, a series of design principles were presented, representing key learnings from the project. The three principles: 1. *Simplify for Scalability and Comparability*, 2. *Ensure Transparency of Evaluation Foundations*, 3. *Prioritize Customizability in Evaluation Tools*, each exploring different dynamics of designing client evaluation frameworks in cooperation with practitioners and end-users. An overall key takeaway is that the analysis should be transparent, so that stakeholders can challenge the assumptions made, and that great value can come from relatively small amounts of data.

Being an ADR paper, a great deal of effort was put into making sure that the paper authentically represented the design process, in alignment with critical research evaluation criteria (Sein et al., 2011). In general, some of the findings in the paper aligned with the findings of other people (Kembro et al., 2018; Thiess and Müller, 2020), but, as researchers, we recognize that the limited setting is unlikely to accurately reflect all 3PL providers.

However, documenting the design of a multi-criteria evaluation model to screen warehouse clients has proven valuable to the industry partner, offering a practical tool for addressing real-world challenges. More importantly, this work exemplifies how collaborative efforts can help bridge the gap between academic research and industrial practice within the 3PL warehousing logistics industry. By contributing to the advancement of multi-criteria evaluation theory and demonstrating its application in a 3PL warehouse setting, this paper not only adds to academia, but also provides actionable insights for practitioners striving to optimize warehouse operations in an increasingly data-based and complex landscape.

References

M. M. Akarte and B. Ravi. Casting product-process-producer compatibility evaluation and improvement. *International Journal of Production Research*, 45(21):4917–4936, 11 2007. ISSN 00207543. http://dx.doi.org/10.1080/00207540600887661.

P. Baker and M. Canessa. Warehouse design: A structured approach. *European Journal of Operational Research*, 193(2):425–436, 3 2009. ISSN 03772217. http://dx.doi.org/10.1016/j.ejor.2007.11.045.

G. Baruffaldi, R. Accorsi, R. Manzini, and E. Ferrari. Warehousing process performance improvement: a tailored framework for 3PL. *Business Process Management Journal*, 26(6):1619–1641, 11 2020. ISSN 14637154. http://dx.doi.org/10.1108/BPMJ-03-2019-0120.

N. Cross. *Engineering design methods: Strategies for product design*, volume 16. 2000.

I. Dahlberg. Conceptual compatibility of ordering systems. *KNOWLEDGE ORGANIZATION*, 10(1):5–8, 1983. ISSN 0943-7444. http://dx.doi.org/10.5771/0943-7444-1983-1-5.

Dataiku Knowledge Base. Concept | Time series forecasting, 12 2024. URL https://knowledge.dataiku.com/latest/ml-analytics/time-series/concept-time-series-forecasting.html.

R. de Koster, T. Le-Duc, and K. J. Roodbergen. Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, 182(2):481–501, 10 2007. ISSN 03772217. http://dx.doi.org/10.1016/j.ejor.2006.07.009.

J. Gu, M. Goetschalckx, and L. F. McGinnis. *Research on warehouse design and performance evaluation: A comprehensive review*, 6 2005. ISSN 03772217.

J. Gu, M. Goetschalckx, and L. F. McGinnis. Research on warehouse operation: A comprehensive review. *European Journal of Operational Research*, 177(1):1–21, 2 2007. ISSN 03772217. http://dx.doi.org/10.1016/j.ejor.2006.02.025.

IBM. *Forecasting requirements*, 3 2024. URL https://www.ibm.com/docs/en/planning-analytics/2.0.0?topic=forecasting-requirements.

Interaction Design Foundation. *KISS (Keep it Simple, Stupid) - A Design Principle*, 2 2024. URL https://www.interaction-design.org/literature/article/kiss-keep-it-simple-stupid-a-design-principle.

D. Jothimani and S. P. Sarmah. Supply chain performance measurement for third party logistics. *Benchmarking*, 21(6):944–963, 1 2014. ISSN 14635771. http://dx.doi.org/10.1108/BIJ-09-2012-0064.

H. Jung. Evaluation of third party logistics providers considering social sustainability. *Sustainability (Switzerland)*, 9(5), 2017. ISSN 20711050. http://dx.doi.org/10.3390/su9050777.

J. H. Kembro, A. Norrman, and E. Eriksson. *Adapting warehouse operations and design to omni-channel logistics: A literature review and research agenda*, 9 2018. ISSN 09600035.

M. Khakdaman, J. Rezaei, and L. Tavasszy. On the Drivers of Demand for Innovative Freight Transportation Services. *IEEE Engineering Management Review*, 51(1):238–249, 2023. ISSN 19374178. http://dx.doi.org/10.1109/EMR.2022.3223313.

G. McGrath. *Data-driven design for sustainable competitive advantage | Geoff McGrath | TEDxGlasgowSalon*, 2022. URL https://www.youtube.com/watch?v=B1Q6xMdMSsU.

T. M. Munsberg, L. Hvam, and L. Tsintzou. *Optimizing the integration process of new customers in a 3PL warehouse*. 2022.

G. Prockl, A. Pflaum, and H. Kotzab. 3PL factories or lernstatts? Value-creation models for 3PL service providers. *International Journal of Physical Distribution and Logistics Management*, 42(6):544–561, 7 2012. ISSN 09600035. http://dx.doi.org/10.1108/09600031211250587.

L. Sebastian-Coleman. The Data Challenge: The Mechanics of Meaning. In *Meeting the Challenges of Data Quality Management*, pages 69–92. Elsevier, 2022. http://dx.doi.org/10.1016/b978-0-12-821737-5.00004-3.

M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. ACTION DESIGN RESEARCH. *Technical Report* 1, 2011.

E. M. Strøm, T. M. Munsberg, and L. Hvam. Identifying Potential Applications of Service Configuration Systems in a Logistics Company. *Technical report*, 2023a.

M. Strøm, E., M. Munsberg, T., and L. Hvam. *Definition & Categorization of Value-Added Services Using a Platform Approach in a Logistics Company*. 2023b.

T. Thiess and O. Muller. Towards Design Principles for Data-Driven Decision Making-An Action Design Research Project in the Maritime Industry. *Technical report*, 2018. URL https://aisel.aisnet.org/ecis2018_rp/144.

T. Thiess and O. Muller. *Setting Sail for Data-Driven Decision-Making an Action Design Research Case from the Maritime Industry*. pages 291–317. 2020. http://dx.doi.org/10.1007/978-3-030-46781-4{_}12.

D. L. Thurston. *Multiattribute Utility Analysis in Design Management*. Technical Report 4, 1990.

Y. Xiao, Y. Cui, N. Raut, J. H. Januar, J. Koskinen, N. Contractor, W. Chen, and Z. Sha. Information Retrieval and Survey Design for Two-Stage Customer Preference Modeling. *In Proceedings of the Design Society*, volume 2, pages 811–820. Cambridge University Press, 5 2022. http://dx.doi.org/10.1017/pds.2022.83.

G. Xiu and X. Chen. The third party logistics supplier selection and evaluation. *Journal of Software*, 7(8): 1783–1790, 2012. ISSN 1796217X. http://dx.doi.org/10.4304/jsw.7.8.1783-1790.