A New Method of Measuring Overall Warehouse Performance: An Automated E-Commerce Retail Warehouse

Student Name: Obaadah Mekhallalati

Student Number: 5163943

Chair: Jafar Rezaei

First Supervisor: Stefano Fazi

Second Supervisor: Zenlin Roosenboom-Kwee

Company Supervisor: Maartje Wammes

A New Method of Measuring Overall Warehouse Performance: An Automated E-Commerce Retail Warehouse

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Obaadah Mekhallalati

Student number: 5163943

To be defended in public on August 26th, 2022

Graduation committee

Chairperson: Dr. J. (Jafar) Rezaei, Section Transport and Logistics

First Supervisor: Dr. S. (Stefano) Fazi, Section Transport and Logistics

Second Supervisor: Dr.ir. Z. (Zenlin) Roosenboom-Kwee, Section Economics of Technology and Innovation

External Supervisor: M. (Maartje) Wammes, CEVA Logistics/Business Analytics

Acknowledgments

I could not have been writing this without the tremendous, continuous, and truly life-changing support that I received from my parents. I remember when I called you in my hardest times and got nothing but the comfort and warmth I was seeking. The countless hours and sleepless nights that you have gone through made me who I am today. I am truly grateful to Allah for blessing me with such amazing parents. My father, Mejahed Mekhallalati, was the one that said the right words at the right times. Baba, you inspired me and was the best living role model that any man would dream of. Thank you for being the shoulder that I cried on and for pulling me out of the slumps with your caring arms whenever I got overwhelmed in this journey. My mother, Hiba Hijazi, was the hero in the shadows honestly. Waking up past midnight to answer my calls and keeping me company in my hardest times. Mama, you are the best listener out there, thank you for keeping up with me when I needed you the most. My parents, thank you for teaching me the valuable life lessons and for providing me with the best education in all my life stages.

My thanks are extended to my siblings as well. Husam, you are not just my brother, but you are my best friend. Thank you for supporting me and telling me the bitter truth when I needed it. I could not ask for a better brother than you and I truly appreciate you standing by my side all the time. Escelle, you probably did not expect this, but I am not exaggerating if I said that some words that you said to me gave me motivation for weeks. I remember when you picked up the phone instead of Mama sometimes and you'd just magically put a smile on my face. And yes, I will mention your cats as well, they also were cute and showed love, especially Dusty. Big thanks to my friends, Azizbek, Musab, Reyhan, Arish, Mohammed & Hamza, and many others. Thank you for cheering me up and special thanks for being my family when living abroad.

This thesis was without a doubt the greatest learning experience I had throughout my studies. I always wanted to do my thesis at a company to be able to use scientific knowledge in practice and bring my work to life. One cannot thank the people who supported me in this journey enough. I would like to thank my committee for their support throughout this project. I would like to thank Maartje Wammes for being a very supportive and particularly cheerful supervisor at the company. My thanks are extended to CEVA logistics for offering me the chance to conduct my thesis within the company. I would like to thank Stefano Fazi for the encouraging remarks and responsiveness throughout this journey, even at very abnormal times, thanks a lot Stefano. I remember when you read my documents at very early mornings when I needed your input rapidly. Jafar Rezaei and Zenlin Kwee have provided more care than I expected as a chairman and second supervisor, they steered me in the right direction many times and gave detailed feedback that was crucial to completing the project. Their words were chosen precisely and encouraged me to improve my work rather than being discouraged that it wasn't complete.

Dedicated to my parents,

Sincerely Yours,

Obaadah Mekhallalati, Eindhoven, 10th of August 2022

Executive Summary

With the ever-lasting expansion of e-commerce retail, there has been less need for retail facilities that customers visit. Instead, many products are sold online and are delivered to the customer straight from warehouses. More warehouses are being built and the efficiency of the operations that occur in these facilities is becoming of higher importance to meet the promises of next-day delivery and in some cases, same-day delivery. Key performance indicators (KPIs) are used to identify the performance of different warehousing operations. The vastness of these KPIs can be overwhelming to some personnel and it can be difficult to keep track of all 130 KPIs that the literature identifies. Therefore, an overall metric that can measure the performance of the warehouse with one meaningful KPI is desired in the warehouse industry.

To ensure the relevance of this research and to ensure its usefulness, a case study with an industry leader took place. One of the world's leading logistical service providers, CEVA, operates numerous warehouses throughout the globe. The case study performed in this thesis was within a warehouse operated by CEVA and owned by one of Europe's largest online fashion retailers. Naturally, there are several technical stakeholders in a warehouse that must work together to ensure smooth warehousing operations. The challenge of finding a way to measure overall warehouse performance (OWP) that addresses the various stakeholders was evident. Furthermore, one problem that faced the case company was justifying KPIs falling short due to shortcomings from the other parties. An OWP measure that addressed the impact of different downtimes on the overall performance of the warehouse was needed.

This thesis aimed to explore a new method to measure OWP. Currently, most studies performed seeking an OWP measure are done using data envelopment analysis (DEA). However, it is difficult to find empirical evidence that DEA has significantly improved performance evaluation and benchmarking in actual non-production conditions. Production theory concepts such as returns to scale and production possibility set may be difficult to apply in pure multiple-criteria evaluation situations, which is why DEA may be less effective in these situations. As a result, this thesis aimed to explore a new way of measuring OWP inspired by the manufacturing industry. One metric that has been widely accepted for many years in manufacturing is called overall equipment effectiveness (OEE). This overall metric has been favored by many managers due to its simplicity and straightforwardness. Moreover, its ability to provide insight to potential areas of improvement makes it more than just a number, hence making it a meaningful KPI. So far, there has been little to no research about using OEE in warehouses which makes this project a pioneer in that direction.

The research done followed a design science research (DSR) approach. This methodology is a paradigm for problem-solving that aims to advance human knowledge through producing original artefacts to solve real life problems. The problem here lied in developing a framework to measure the OWP of an automated warehouse using a modified OEE. Firstly, a modified OEE framework was developed. Numerous warehouse tours took place, along with interviews with personnel from different organizational functions to identify key KPIs and critical success factors (CSFs) that will shape the modified OEE framework. The results of these interviews and tours were a deeper understanding of the warehouse operations and what should be considered in an OEE model for warehouses.

An artefact in the form of an Excel model was built and evaluated by experts in the field qualitatively. Semi-structured interviews took place to leave room to discuss issues that may arise during the interview. The artefact was evaluated for its complexity, applicability in practice, relevance to industry, and its utility. Interviewees were given room for suggestions to improve the artefact and concerns about the model. After that, a quantitative evaluation of the model took place. A sensitivity analysis to ensure the OEE changes in relatively similar increments to input changes was done. Moreover, a JASP Bayesian correlation test was done to test if using OEE as an OWP is truly representative of the performance. OEE value trends were compared to score values of what was described as the singular most crucial outcome score for inbound, namely dock-to-stock time. Although the results do not support a positive correlation between OEE and dock-to-stock scores, further research to validate this more is needed to provide insight on whether OEE is truly representative of OWP.

Throughout the research and after interpreting the results and analyzing data, several practical and theoretical implications were noted. It was interesting to observe different mentioned potential uses for the artefact by different employees, showcasing a diverse range of utilities of OEE in warehousing. Moreover, the single use of the model that was agreed upon by all participants was reporting this measure to different stakeholders. This aligns with the purpose of having an OWP measure, simplicity. Furthermore, using the tool to identify improvement areas gives OEE the ability to steer continuous improvement within a warehouse. Theoretical implications arose from observing a relationship between OEE coefficients. While they are stand-alone factors by definition, there is a hidden impact of downtime on performance efficiency that was noted. This challenges the interdependence of OEE coefficients. In addition to that, the use of weights to magnify certain coefficients in OEE calculations may cause quality and performance tradeoffs. However, in practice, minimum thresholds above 99% for quality KPIs ensure that this is not exploited to arrive at a higher OWP measure. Recommendations of reevaluating target KPIs were made to arrive at more realistic OEE values. Moreover, guidelines to using the model as a comparison tool between warehouses was discussed since the use of OEE as a benchmark has been challenged by many.

As the research regarding the use of OEE in warehousing is scarce, this project serves as a starting point for exploring the use of OEE in warehouses. The study employed a single case study in the inbound section of an automated warehouse for online clothes retail. Many more goods and warehouses may necessitate alternative definitions of warehouse losses and OEE factors. It is advised that this study be expanded into additional warehouse areas, and that a final OWP employing a modified OEE model be created after multiple iterations and case studies. It is believed that by employing OEE, a universal indicator of some kind can be established in the future, which can then be used for benchmarking similar warehouses. Finally, after establishing the usefulness of OEE as an OWP metric, further quantitative analysis of its performance compared to other commonly used approaches such as DEA should be included in future study.

List of Abbreviations

Distribution Center (DC)

Supply Chain (SC)

Information Technology (IT)

Information Systems (IS)

Overall Equipment Effectiveness (OEE)

Total Productive Maintenance (TPM)

Supply Chain Management (SCM)

Active Research (AR)

Key Performance Indicators (KPIs)

Data Envelopment Analysis (DEA)

Constant-return-to-scale (CRS)

Variable-return-to-scale (VRS)

Analytic hierarchy process (AHP)

Critical success factors (CSFs)

Production Equipment Effectiveness (PEE)

Overall Factory Effectiveness (OFE)

Overall Warehouse Performance (OWP)

Design Science Research (DSR)

Design Science Research Methodology (DSRM)

Service Level Agreements (SLAs)

Seconds per Item (SPI)

Warehouse Management System (WMS)

Decision Making Unit (DMU)

Table of Contents

Acknowledgments....3
Executive Summary....4
List of Abbreviations....6
Table of Contents....7
List of Figures....11
List of Tables....11
Chapter 1: Introduction....12
1.1 Background.... 12
1.2 Research Relevance.... 15
1.2.1 Problem Definition.... 15
1.2.2 Research Objective.... 16
1.2.3 Research Question.... 16
1.2.4 Practical and Societal Relevance.... 16
1.2.5 Brief Summary of the Case Study.... 17
Chapter 2: Literature Review....18
2.1 Warehouses in Literature.... 18
2.1.1 Warehouse Types.... 18
2.1.2 Warehousing Activities.... 19
2.1.3 Warehouse Performance Metrics.... 20
2.2 OEE in Literature.... 26
2.2.1 OEE Background.... 26
2.2.2 OEE Evolution.... 28
Chapter 3: Methodology....33
3.1 Research Design.... 33
3.1.1 DSR vs AR.... 33
3.1.2 DSR Definition and Objectives.... 35
3.1.3 Implementing DSR.... 36
3.1.4 DSR Evaluation.... 39
3.1.5 Design Knowledge Contribution.... 40
3.2 Data Collection and Analysis.... 41

3.2.1 Unstructured and Semi-structured Interviews 41
3.2.2 Informal Methods 43
3.2.3 Data Analysis 43
3.3 Research Validity 44
Chapter 4: Artefact Development 45
4.1 Identifying OWP Factors 45
4.1.1 Stakeholders in a Warehouse 45
4.1.2 Important KPIs 46
4.1.3 Critical Success Factors at the Case Warehouse 46
4.2 Modifying OEE for Warehouses 46
4.2.1 Warehouse Losses 47
4.2.2 Modified OEE Framework 48
4.3 Building the Artefact 50
4.3.1 Availability 50
4.3.2 Performance 51
4.3.3 Quality and OEE Calculation 52
4.4 Applications of OEE in Warehousing 53
4.4.1 Tracking Performance 53
4.4.2 Continuous Improvement 53
4.4.3 Communicating OWP Results 54
4.4.4 OEE Benchmark and Thresholds 54
Chapter 5: Evaluation 56
5.1 Complexity of the Artefact 56
5.1.1 Initial Impressions 56
5.1.2 Difficulty of Use and Familiarity 56
5.1.3 Possible Confusions and Final Impressions 57
5.2 Applicability and Relevance of the Artefact 57
5.2.1 Inputs of the Model 58
5.2.2 OEE Components 58
5.2.3 Matching Expectations 59
5.3 Applications of OEE in Warehousing 59
5.3.1 Reporting to Stakeholders 60

5.3.2 Tracking Performance 60
5.3.3 Continuous Improvement 60
5.3.4 Comparing Performance to Other Warehouses 60
5.4 Suggestions and Concerns 61
5.4.1 Improving User Interface 61
5.4.2 Connecting to System Interactively 61
5.4.3 Input Data 61
5.5 Quantitative Analysis 62
5.5.1 Sensitivity Analysis 62
5.5.2 Correlation Test 64
Chapter 6: Discussion 67
6.1 Evaluation Results 67
6.1.1 Complexity of the Model 67
6.1.2 Applicability and Relevance of the Model 67
6.1.3 Utility of OEE 68
6.1.4 Quantitative Analysis 68
6.2 Practical Implications 69
6.2.1 Addressing Areas of Improvement in the Warehouse 69
6.2.2 Performance and Quality Tradeoff 69
6.2.3 Comparing Warehouse Performance 70
6.3 Scientific Implications 70
6.3.1 Interdependency of OEE Coefficients 70
6.3.2 Using OEE as a Benchmark 72
6.4 Recommendations and Reflections 73
6.4.1 Recommendations for the Industry 73
6.4.2 Limitations and Further Research Directions 74
Chapter 7: Conclusions 75
References 77
Appendices 83
Appendix A 83
Appendix B 84
Appendix C 85

Appendix D 86

List of Figures

Figure 1 Thesis Outline 15
Figure 2 Warehouse Types based on function 19
Figure 3 Automation Levels of Warehouses 19
Figure 4 General Structure of warehouse in practice 20
Figure 5 General OEE framework (TPM view); page 26 (Pintelon et al., 2000) 28
Figure 6 DSRM process framework, Source: (Peffers et al., 2007) 37
Figure 7 DSR knowledge contribution framework, Source: (Gregor & Hevner, 2013) 41
Figure 8 Modified OEE framework for warehouses (inbound) 48
Figure 9 OEE and dock-to-stock scatter plot 66
Figure 10 Hidden impact of downtimes 71

List of Tables

Table 1 KPIs per dimension of measurement, page 9 (Staudt et al., 2015) 23
Table 2 Warehouse key performance indicators (Kusrini et al., 2018) 24
Table 3 List of OEE based models from (Ng Corrales et al., 2020) 30
Table 4 DSR vs AR; page 95-96 (Dresch et al., 2015) 34
Table 5 DSR guidelines, Source: (Hevner et al., 2004) 39
Table 6 DSR evaluation methods, Source: (Hevner et al., 2004) 40
Table 7 DSR knowledge contribution types, Source: (Gregor & Hevner, 2013) 40
Table 8 Interviewees overview 44
Table 9 Availability input in Excel model 51
Table 10 Performance input in Excel model 51
Table 11 Quality input table in Excel 52
Table 12 OEE calculation and weights input in Excel model 53
Table 13 Complexity of artefact evaluation 57
Table 14 Applicability and relevance of artefact evaluation 59
Table 15 Artefact utility evaluation 61
Table 16 Sensitivity analysis of Excel model 63
Table 17 OEE vs Dock-to-stock scores 64
Table 18 Bayesian correlation test results 64
Table 19 Bayesian Factor values interpretation, Source: (Tona et al., 2019) page 8 65
Table 20 Pearson's r interpretation, Source: (Ratnasari et al., 2016) page 2 65

Chapter 1: Introduction

With the current trend in less physical retail stores being deployed and with the rise of home deliveries for consumer products, many items are being shipped to customers straight from a warehouse rather than being displayed in a physical retail store. Warehouses are a central part of any supply chain as they act as the link between the suppliers and the customers. The storage and retrieval of products are essential for manufacturing, logistics, and delivering services (Faveto et al., 2021). The warehouse industry is constantly growing with a global market value of above 450 billion dollars in 2021 and a forecasted growth to 605.6 billion dollars in 2027 according to IMARC Group (Warehousing and Storage Market: Global Industry Trends, 2022).

1.1 Background

E-commerce is increasing demand, it is forecasted that global e-commerce sales are to reach 7.4 trillion dollars by 2025 as opposed to the 5.4 trillion-dollar sales in 2021 (Cramer-Flood, 2022). This rise in online shopping led to an increased demand in warehousing for leading companies and new competitors. The recent development in information technologies (IT) and transportation positively impacts the efficiency of warehouses nowadays (Faveto et al., 2021). In the fiercely competitive market of online retailing, there comes a need for continuous improvement of logistical services and systems. Many stores now offer next day delivery, and some are implementing same day delivery nowadays (Ulmer, 2017). This can only be possible with an efficient supply chain connecting all the stops that an item goes through, from the manufacturing supplier and ending with the item on the customer's doorstep. To ensure such efficiency in supply chains, efficient and effective warehousing activities must be achieved.

The growth of demand globally has changed the way inventory is managed. A trend of reducing inventory by having bigger and fewer distribution centers (DCs) has been observed (van den Berg & Zijm, 1999). For some companies a small number of DCs serve an entire continent or region like Europe or the Middle East. Material handling involves the movement of materials to, through, and from production processes; within warehouses; and in receiving and shipping (van den Berg & Zijm, 1999). Material flow and warehousing are the composing elements of material handling. Various devices are used for material flow. Examples include conveyors and forklifts; these devices facilitate the movement of products within a warehouse. Hence it can be said that warehousing involves all material handling activities withing the receiving and shipping areas of the warehouse.

Naturally, performance measures are vital for any firm that wishes to be competitive. They provide a quantitative measure that can be compared to previous values of the same firm or competitors' values if they are known. Being able to identify on which side of the spectrum a warehouse lies in, is of great importance when assessing the stance of a firm in the market from a performance perspective. Performance measures also indicate where shortcomings happen and can help identify where a problem lies in the system which in turn sheds a light on areas the management should improve. A performance measure must evaluate the current system and provide a ground for future planning based on the warehouse system analysis.

Of course, performance measures differ from one industry to another. In the warehouse business, many different key performance indicators (KPIs) are used. Organizations tend to prioritize some KPIs over others based on their strategy and goals. A warehouse that stores medicines and drugs may prioritize KPIs that ensure quality over a warehouse that stores wooden assembly parts for example. In practice, an overall metric, a number that can summarize the performance of a warehouse is often needed and utilized. An overall metric can indicate if an improvement in a certain KPI has a significant change on the collective performance of the warehouse. Additionally, non-technical personnel that may not know all the detailed KPIs of such a warehouse may have a good idea of the overall performance through a universal metric. This is particularly useful when marketing service providers' services and their effectiveness to potential customers. Other applications are further discussed throughout the report and in the discussion chapter.

While there is no consensus on an overall metric to cover all warehouses, most research done in this area used a method known as data envelopment analysis (DEA) (Faber et al., 2018). However, it is uncommon to find empirical proof that DEA has significantly enhanced performance assessment and benchmarking in actual non-production contexts. The fundamental principles of production theory, such as returns to scale and production possibility set, may be the main cause of this shortcoming in DEA. Pure multiple-criteria assessment issues, which are frequently attempted to be solved with DEA, are difficult to apply these notions to. (Wojcik et al., 2019)

Overall warehouse performance (OWP) has mostly been measured using DEA in the literature. As noted by Wojcik et al. (2019), the use of DEA in non-production contexts such as warehousing seldom improves performance assessment. The main purpose of an OWP is to provide a summarized assessment of a warehouse. Hence it is crucial to use a measure that can not only improve assessment and benchmarking, but also provide insight as to where improvements should be made to increase overall performance. The interaction between KPIs of a warehouse and the impact of downtimes on performance measures needs to be addressed.

In practice, a problem that faces logistical service providers lies in addressing downtimes caused by other business partners in a warehouse. As a service provider relies on resources such as the warehouse management systems (WMS) and mechanical conveyor systems of other stakeholders, it is important to identify the causes of downtime and its impact on performance measures. An overall measure was needed and is expected to consider these interactions and help provide insights on the impact of downtime of different stakeholders of performance measures. To solve this problem, a design science research (DSR) methodology is used. The aim of this methodology is to provide a solution to a real-life problem through the development of an artefact or model. Using this methodology helps escaping the "consultant" role that a researcher may fall into as it emphasizes contributing to the knowledge base as a product of the research. The goal of this research is thus to contribute to solving this problem that faces many service providers through the development of an OWP measure that addresses their needs. An overall metric that can be used as an OWP measure through combining independent factors can provide a more practical and useful application of OWP.

One metric that is becoming increasingly popular as a general performance measure in manufacturing nowadays is Overall Equipment Effectiveness (OEE) (Muchiri & Pintelon, 2008). This metric was first introduced by Nakajima in 1988 as a part of Total Productive Maintenance (TPM) (Nakajima, 1988). OEE was developed to measure the effectiveness of manufacturing systems. In recent years, there has been an increasing focus on OEE, and more papers have been published addressing it. In fact, more than 50% of the publications regarding OEE were published from 2015 onwards, showing an increased interest in the subject (Ng Corrales et al., 2020). Despite the metric being intended for the manufacturing sector, several authors have added other variables to the OEE metric and modified it to suit specific industries. Examples of areas where OEE was used after an adaptation to the industry include sustainability, transport, mining, electricity, and resources. The use of OEE as a performance measure for warehouses or in logistics generally has not been researched extensively yet. Some authors have explored this area and showed how they adapted OEE to logistical case studies. However, the application of OEE within warehouses has not been explored yet.

OEE adaptation presents a hurdle because it was initially created for a single piece of production machinery and has undergone numerous changes since then. The modification of OEE for this industry is a crucial aspect of this research because an adaptation to warehousing has not yet been studied. Therefore, studying the components of such model needs to be done.

In an effort to provide an overall warehouse performance measure for warehouses that can be developed to be a universal indicator; and building upon the rising interest in OEE in the field of logistics, this thesis will aim to contribute in building a framework for a modified OEE for automated warehouses in the online retail industry from a logistics service provider’s perspective which will serve as a contribution to the use of OEE in logistics.

This thesis is outlined in the following way. First, the problem is introduced and a background on its relevance is outlined in the introduction chapter. After that, the literature review chapter addresses the relevant literature found as a basis for conducting this research. Followed by the second chapter is the methodology section. An overview of the chosen methodology and how data is collected and analyzed is discussed in this chapter. In the fourth chapter, the artefact development stages were discussed, and the model is presented. Subsequently, an evaluation chapter shows a qualitative evaluation of different employees to the artefact; additionally, a quantitative analysis of the model is presented. Finally, discussions, implications, recommendations, and conclusions are presented in the final chapter. Figure 1 shows an outline of the project.

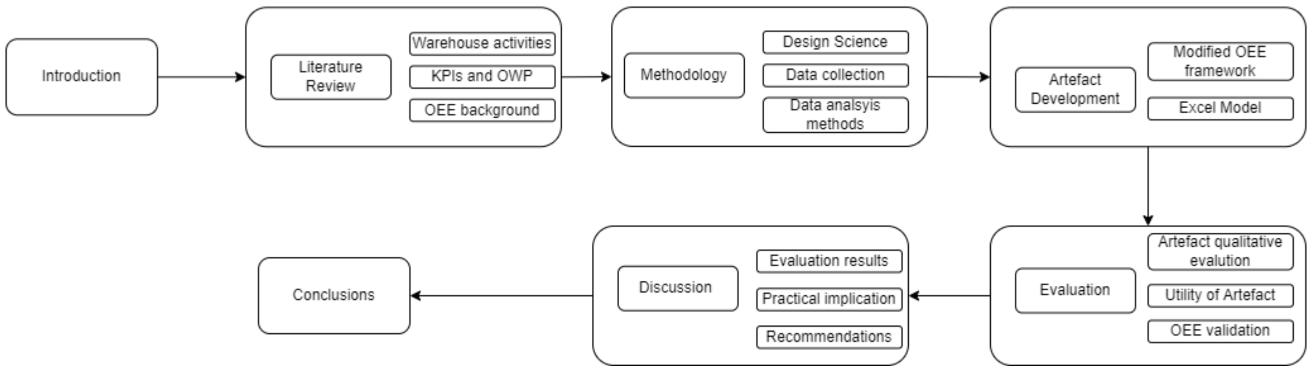

Figure 1 Thesis Outline

1.2 Research Relevance

After exploring the literature regarding warehouses, its types, performance measures, and the categories of these KPIs; and upon the lack of consensus on overall warehouse performance measures; it would be useful to develop a model that may serve as a contribution to the few measures concerning OWP.

One of the widely accepted overall metrics in the manufacturing industry and that are favored by managers due to its conciseness and collectiveness, is what is known as OEE (Ng Corrales et al., 2022). As has been observed, this metric was adopted by different industries and modified by researchers and practitioners to meet their needs. So far, after exploring the literature concerning warehouses and OEE, there has not yet been a contribution to measure OWP using an OEE based approach.

As the scope of research within OEE is expanding, and upon the recommendations of various researchers to expand the research in the direction of OEE and logistics; this research of building a modified OEE framework to measure overall warehouse performance may be a pioneering effort in that area.

1.2.1 Problem Definition

The research problem lies in evaluating the overall performance of warehouses through the use of a modified OEE approach. In order to ship products to customers on time, the performance of these warehouses must be tracked. An accurate examination of the system will also provide a more accurate basis for expectations from the warehouses to deliver to customers on time, reducing the chances of unrealistic and unmet promises. As asserted by Frazelle (2002), the use of case studies to study warehouse performance measures is the most optimal approach. To combine the knowledge in the scientific literature about warehouse KPIs and the OEE evolution with the industrial practice, a case study examining an automated warehouse will take place.

The research is done within a logistics and technology solution providing company, CEVA logistics, that operates the warehouse owned by a large online retailer. The problem CEVA is facing lies in identifying the hidden impact of system downtime on logistical uptime. Tackling this problem through the implementation of a modified OEE framework will provide a science-based ground to indicate or identify the parameters leading to a low measure of OEE from its coefficients or its coefficient components, hence will serve as a foundation for continuous improvement. Using OEE as an OWP for an automated warehouse will provide a unified and objective way of evaluating its system. This evaluation can then be an indicator to which direction improvements can be made to minimize the deviation between expected objectives and actual results.

1.2.2 Research Objective

To address the gap in the literature with regards to a modified OEE in the logistic sector and to provide a solution to the problem stated, the following objective is proposed.

Develop a framework to measure the OWP of an automated warehouse using a modified OEE

To achieve this objective, the following sub-objectives must be met:

- Identify the current situation in the literature for OEE in logistics
- Develop a modified OEE formula for an automated warehouse
- Build a framework that addresses the real impact of system downtime on target KPIs

1.2.3 Research Question

To achieve the mentioned objectives, the following research questions must be answered.

Main Research Question:

How can we modify and utilize an OEE framework for an automated warehouse to measure its overall warehouse performance?

Sub questions:

1. What warehouse KPIs are relevant for such overall metric and what are the critical success factors (CSFs) of the warehouse in study?
2. What OEE frameworks for a logistical process are present in the current literature and how was OEE modified in this field?
3. How can the impact of subsystems downtime on the system downtime and logistical uptime be measured using a modified OEE?
4. How can a modified OEE framework serve as a science-based ground for continuous improvement?

1.2.4 Practical and Societal Relevance

The implications of this research are not limited to expanding knowledge in the use of OEE for warehousing. In fact, the purpose of advancing knowledge can often be improving life quality and easing the operations. Providing a one stop station for evaluating warehouse performance can be beneficial to the consumer receiving the products. Monitoring performance to identify areas of improvement in warehousing operations will directly impact the consumer receiving the shipments. Having reliable delivery times and met promises encourages customers to continue purchasing online. This can help increase the overall flow of the economy as more people are able to purchase their needs and wants with no problems.

Effective warehousing means parcels delivered on time, undamaged and at a higher volume. This does not only mean more profits for companies, but also means an increased dependency on online shopping for customers. This leads to more spaces that can be utilized more efficiently withing cities rather than storing products on shelves for long periods of time. It is argued that an OWP measure using OEE can be a potential to eventually identify areas of improvement to increase warehousing effectiveness.

1.2.5 Brief Summary of the Case Study

CEVA is a supply chain technology and service provider serving many clients, one of which is a major online retail company where the case study takes place. As this retailer aims to widen its logistics network to provide products to more customers and provide them swiftly, it launched a new warehouse that will help facilitate next and same day deliveries. CEVA, which is acting as the logistical solution provider, wants to build a framework that can measure a modified OEE for a logistical process. More specifically, the aim of this investigation is to identify the actual impact of the mechanical and warehouse management systems’ downtime on the logistical uptime of the retailer. Showing the hidden losses and downtimes and identifying the most impactful processes on the overall logistical uptime. This framework will help CEVA to report the key areas that need to be improved in the system. The framework will also provide a science-based foundation for continuous improvement at the warehouse and other similar systems making CEVA a leader in this venue.

Chapter 2: Literature Review

Before diving into research, one must explore the current literature regarding the scope of study. This is done to ensure originality of the research and to identify potential papers that can be built upon to expand the research.

For this research, Scopus was used to search for papers. Primarily Scopus was used to obtain relevant papers due to its increased functionality and ease of use. The keywords “warehouse” and “OEE” were first used to see if there are currently any papers that discuss the use of OEE in warehousing. This resulted in 10 papers, however, upon scanning the titles and abstracts none of them discuss the use of OEE as an OWP measure. This resulted in the need to search for papers that address warehouse performance, KPIs, and OWP. In addition to that, a search was done for papers discussing OEE evolution. A systematic literature review for warehouse KPIs by Staudt et al. (2015) served as a key paper in the field of warehouse performance evaluation for this project. Additionally, a thorough systematic review by Ng Corrales et al. (2020) served as a key paper for the OEE side of this literature review. A snowball approach followed to arrive at a collection of papers relevant to the project. Abstracts were scanned to ensure that a paper is to be included. Inclusion criteria included papers that described warehouse activities and different KPIs for each of these activities; KPI classifications and OWP approaches were also to be included for this review. Inclusion criteria for OEE papers mainly ensured the search for papers that showcase the evolution of OEE. In addition to the paper written by Ng et al., another comprehensive OEE review by Muchiri & Pintelon (2008) served as a key paper for this review.

In this chapter, firstly warehouses, warehousing activities, and KPIs will be discussed. After that, an OEE background is provided, and its evolution is outlined as well as its applications in different industries.

2.1 Warehouses in Literature

2.1.1 Warehouse Types

Van den Berg & Zijm (1999) distinguish warehouse types by dividing them into 3 categories (Figure 2) based on their purpose. Firstly, distribution warehouses where products from different suppliers are collected to then be delivered to various customers. Secondly, production warehouses are used for storing raw materials, semi-finished products, and finished ones. This type of warehouse often is in conjunction with the producing factory rather than storing for multiple suppliers. Finally, contract warehouses are where the warehousing operations are done on behalf of a customer.

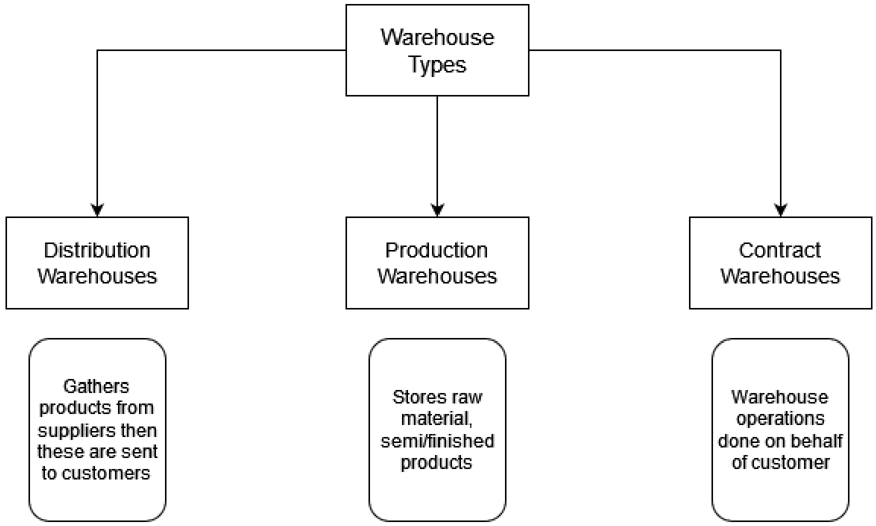

Figure 2 Warehouse Types based on function

Another classification of warehouses is based on the level of automation used as shown in Figure 3. A manual warehouse is where all material handling activities are done by manual labour and no automation is present. Automated warehouses are warehouses that have some level of automation but still requires a worker to perform the picking of products. Automatic warehouses require very minimal human involvement and are fully automatic including picking (Faveto et al., 2021).

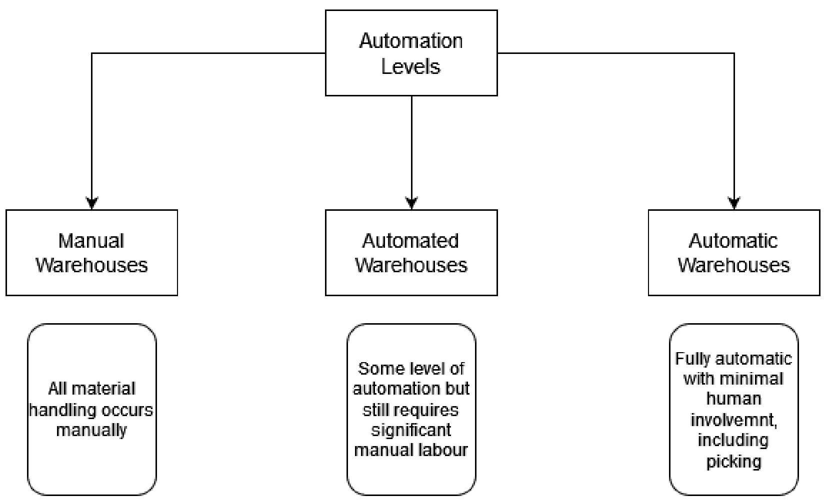

Figure 3 Automation Levels of Warehouses

2.1.2 Warehousing Activities

When studying literature in warehousing, it is clear that there are four main warehouse activities that all warehouses undergo. These activities outline the general operations of warehouses.

Starting with *Receiving* which involves trucks being assigned to docks and then being unloaded to the warehouse. This is then followed by what is known as *Storage*, this is where products are transferred from the unloading area towards their assigned place in inventory or storage. After that, *Order Picking* takes place. This is the area of the warehouse that involves the most labour activity. It marks the beginning of order preparation where the products are picked from the shelves and sent towards their next stop. Then comes the actual *Shipping*, it involves packing the orders and loading them in the assigned trucks. Assigning delivery trucks to outbound docks is also part of this activity.

Some include a fifth activity which would be *Delivery*, as the name suggests, it is the transit of the order of products from the warehouse to the customer (Staudt et al., 2015).

However, in practice, warehouses may expand one of these activities into two areas. Although the general theme of warehouses follows the aforementioned activities, there may be some confusion if more activities take place or if different terminologies were used than the standard. Nevertheless, the definitions of these activities clarify the distinction between them. Figure 4 depicts the general flow of products in one of the largest online retailers' warehouses. It can be noted that although the terms may be different, the areas of the warehouse follow the general definitions mentioned. This division of activities into 6 areas is very similar to the Frazelle classification of dividing warehouse activities into (Receiving, Put-away, Storage, Order Picking, and Shipping) (Frazelle, 2002)

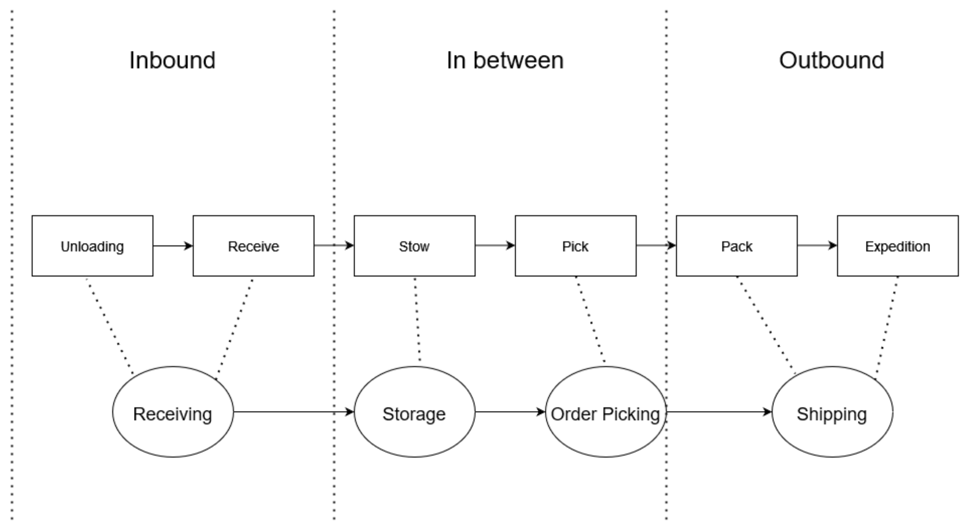

Figure 4 General Structure of warehouse in practice

As shown in the figure above, there is a clear distinction between the four activities, but the defining limits of inbound and outbound activities in a warehouse are not as clear in distinction. Stowing and Picking can be regarded as a "buffer" zone as some experts describe it with the storage activities leaning more towards inbound and order picking being considered the beginning of outbound processes.

2.1.3 Warehouse Performance Metrics

The increasing complexity of warehouses and their activities nowadays with the development of new technologies and techniques makes performance tracking of greater importance (Wu & Dong, 2008). This entails a challenge to warehouse managers to keep track of the right indicators that are most relevant to the organization's strategy and goals. Warehouse KPIs have been explored by researchers in different ways. The diversity in their approaches is due to many factors. These include the objectives to be achieved by the warehouse management, measurement methods, different types of warehouse systems, and the focus area inside the warehouse (e.g. receive and picking) (Staudt et al., 2015). There is no general agreement on a group of metrics to assess a warehouse, hence it is not an easy task to pick amongst many indicators that are found in the literature or in practice.

In order to come up with an overall metric for warehouses while there is no general consensus on the indicators, one must study these indicators and have a greater understanding of what they measure, how often they are used and why some are used over other metrics. The first article to extensively study short-term warehouse performance measures was written by Staudt et al. in 2015. This article will serve as a key paper in this research as it reviews 43 articles that discuss warehouse performance measures. As mentioned earlier, there are five main warehouse activities (including Delivery). Each of these activities may have KPIs specific for them only. According to the authors, the majority of the articles studied included order picking in their study as it is responsible for 60% of the costs of warehouses (Manikas & Terry, 2009).

In their paper, the authors discuss operational warehouse performance measures as opposed to warehouse design and location for example. These operational warehouse KPIs are evaluated and grouped up in categories as will be discussed further. To clarify what is meant by performance analysis, the authors used the following definition: 'the periodic measurement and comparison of actual levels of achievement with specific objectives, measuring the efficiency and the outcome of corporation' from (Lu & Yang, 2010). This narrows down the scope of research and in turn provides more accurate results as it is specific to operational performance measures.

Other researchers have also discussed warehouse performance metrics. In his book, Frazelle describes warehouse performance indicators related to each of the warehouse areas from different perspectives. While his categorization of the metrics was different to that of other researchers, there often is an overlap between the KPIs, but a different way of showcasing them.

2.1.3.1 Categorizing Warehouse Performance Metrics

Key performance indicators (KPIs) in logistics are often referred to as either "hard" or "soft" metrics. Hard metrics are the product of calculations and mathematical expressions and are expressed quantitively. Soft metrics deal with qualitative measures like customer satisfaction which cannot be easily measured (Chow et al., 1994). Following what the previous literature describes them, the hard metrics will be referred to as direct metrics and soft metrics will be referred to as indirect metrics.

Most research that has been done on warehouse performance measures discusses the direct metrics. Examples include labour productivity and put-away time. These are easily calculated, and their parameters are easily measured and incorporated in straight-forward formulas. Moonen et al. (2005) have counted up to 130 metrics, among them are short-term and long-term metrics.

When looking at warehouse performance literature, although it is not as abundant as other areas of literature, one may notice a trend of categorizing KPIs on certain grounds. Some categorize direct KPIs based on the area of the warehouse. Hence *Receive* has its own KPIs and *Picking* has its other separate metrics. Other researchers mention the main dimensions being measured, and categorize the metrics based on that; while another method of grouping KPIs is by the goal they achieve, such as productivity or financial goals. Finally, an approach of measuring the technical efficiency of changing inputs of the warehouse into outputs has been used by others. While each approach has its own strengths, using multiple grouping methods together may result in more clear and accurate listing of important KPIs.

Direct metrics of warehouse performance indicators usually fall into one of four dimensions as stated by Staudt et al. in their paper. These dimensions are *time, quality, cost, and productivity*. Table 1 from their paper can clearly show the how this categorization differentiates between KPIs studied in 43 publications.

Dimension	**KPI**	**Frequency in Literature Studied**
Time	Order lead time	9
	Receiving time	5
	Order picking time	4
	Delivery Lead time	3
	Queuing time	2
	Put away time	2
	Shipping time	2
	Dock-to-stock time	2
	Equipment downtime	1
Quality	On-time delivery	10
	Customer Satisfaction	8
	Order fill rate	5
	Physical inventory accuracy	5
	Stock-out rate	4
	Storage accuracy	4
	Picking accuracy	3
	Shipping accuracy	2
	Delivery accuracy	2
	Perfect orders	2
	Scrap rate	2
	Orders shipped on time	1
	Cargo damage rate	1
Cost	Inventory cost	7
	Order processing cost	3
	Cost as a % of sales	3
	Labour cost	2

	Distribution cost	2
	Maintenance cost	2
Productivity	Labour productivity	11
	Throughput	10
	Shipping productivity	7
	Transport utilization	5
	Warehouse utilization	4
	Picking productivity	3
	Inventory space utilization	3
	Outbound space utilization	3
	Receiving productivity	2
	Turnover	2

Table 1 KPIs per dimension of measurement, page 9 (Staudt et al., 2015)

It can be noted that the number of KPIs included in this table does not add up to the 130 metrics that Moonen et al. studied. The reason is that some metrics may have different names but measure the same thing, and vice versa. It may have the same name but refer to something slightly different than a similar KPI. This was overcome by the authors of this table by eliminating duplicates and ensuring that all metrics are included. However, some have criticized this list and showed that it does not include all the inputs of a warehouse and its activities and the consequent outputs (Karim et al., 2021). Hence, they have revised the paradigm proposed by Staudt et al. and considered the measurement of all inputs and outputs in terms of technical efficiency.

Other warehouse KPIs that have been developed recently and are gaining more attention are KPIs related to warehouse sustainability. Sustainability evaluation is a critical work that is desperately needed; because of the diversity of sustainability and the difficulty of implementing and quantifying sustainability in warehouses, policymakers face several hurdles (Torabizadeh et al., 2020). Examples include: the ratio of recyclable material used to total material used, amount of reduction of energy consumption, extent of decreasing the impact of the organization on biodiversity and habitat, and the percentage of product sold with recycle packaging material. These KPIs are part of an extensive table in a study about sustainable warehousing. Torabizadeh et al. (2020) used the most relevant literature on performance assessment to create this pool of indicators.

In a recent study, new warehouse KPIs were developed in a case study for a textile company (Costa et al., 2023). A list of newly defined KPIs was developed concerning different dimensions. A first expire first out policy at the case company called for a KPI ensuring the policy is being adhered to. Another KPI was regarding storage, a ratio of goods sold to the number of products on shelves in stock was considered a storage KPI. Other KPIs included inventory per square meter and the average stock of each product. A performance monitoring system was developed by the authors of this study and an emphasis on visualization of KPI results was made.

In another case study using the Frazelle classification, 25 performance measures were given to each area of the warehouse and at the same time to which objective of the organization this KPI aims to achieve. Table 2 shows how the combined categorization based on area of warehouse and based on the goals being achieved was presented together.

	Financial	Productivity	Utilization	Quality	Cycle Time
Receiving	Receiving cost per line	Receipts per man-hour	% Dock door utilization	% Receipts processed accurately	Receipt processing time per receipts
Put away	Put away cost per line	Put aways per man-hour	% Utilization of put away labour and equipment	% Perfect put aways	Put aways cycle time (per put away)
Storage	Storage space per item	Inventory per square foot	% Locations and cube occupied	% Locations without inventory discrepancies	Inventory days on hand
Order Picking	Picking cost per order line	Order lines picked per man-hour	% Utilization of picking labour and equipment	% Perfect picking lines	Order picking cycle time (per order)
Shipping	Shipping cost per customer order	Orders prepared for shipment per man-hour	% Utilization of shipping docks	% Perfect shipments	Warehouse order cycle time

Table 2 Warehouse key performance indicators (Kusrini et al., 2018)

Other categories were prescribed by Moonen et al. (2005) that describe warehouse performance metrics in four directions. They stated that the 130 metrics fall in these four categories, namely, *Effectiveness, Efficiency, IT & Innovation,* and *Satisfaction.* This is different to the achieved goals being measured in Table 2 for example. Although some may overlap, for example a larger value for the percentage of perfect shipments may be considered a metric in the satisfaction category, the approach of classifying KPIs is different and may suit one stakeholder’s goals more than the other. Perhaps it may be easier to choose from KPIs related to each warehouse area for an organization as opposed to picking KPIs that are categorized based on the dimension they measure.

2.1.3.2 Overall Metrics for Warehouses

As can be observed from Table 1, some indicators clearly stand out based on the frequency they were noted in the literature; not only that, but also there is no universal indicator for warehouses in general, nor is there a universal indicator to combine the indicators of each dimension. In Table 2, there were 25 KPIs for each goal and area, however they were not summed up in one overall metric. There have been efforts in studying the overall performance of warehouses or prioritizing certain KPIs using different methods. The abundance of metrics is often useful for detailed analysis of the current situation, but a “summarized” version would be easily interpreted by non-professionals in the warehouse specifics and in its departments. This can be useful when reporting to experts from other fields that want to acquire more information about a specific warehouse, like higher management of a logistics firm, or a third-party company that wants to acquire the warehouse or even the contracting partner of a warehouse. Hence an overall metric that serves this purpose is of a great importance for anyone who wants to know the performance of a warehouse at a glance.

The overall warehouse performance has been mentioned in publications that discuss warehouse KPIs. In some instances, it was referred to as the overall technical efficiency, overall warehouse performance, overall efficiency, and overall warehouse productivity. While different methods were used to judge this overall performance, they all had the aim of presenting the warehouse performance in simpler terms.

Most of the research done in this direction used Data Envelopment Analysis (DEA) to come across a single score of warehouse performance. DEA is a non-parametric linear programming method that converts all the data of warehouse inputs and outputs into a single score of performance. The relative efficiency (performance) of a group of comparable decision-making units (DMU) is measured by DEA, where the DMUs are the organizations under examination such as warehouses (Faber et al., 2018). DEA has been used to provide management with estimates of the warehouse efficiency. In their study they used a constant-return-to-scale (CRS) approach for their model, however in other research a variable-return-to-scale (VRS) DEA model was used (Johnson & McGinnis, 2011). They utilized a VRS model with an input orientation that forces warehouses to benchmark output levels against like sized warehouses. Management can utilize the efficiency estimates to compare warehouse performance to a benchmark warehouse, which is a convex combination of observed warehouses.

Often in calculating warehouse performance, intensity weights or importance factors are given to different metrics on various grounds. Some KPIs are prioritized for their importance to the management, other times these are ranked using the analytic hierarchy process (AHP) as was done in (Kusrini et al., 2018).

Other methods such as using a modified grey model has been used by (Islam et al., 2021). They used the widely accepted Frazelle classification of dividing the KPIs into five main categories: financial, productivity, utilization, quality, and time. The AHP ranking method was used alongside the expert opinion on what KPIs to be used. Frazelle contended that using case studies was the best method for developing warehouse benchmark performance indicators. This method of including expert’s opinions on KPIs that match the critical success factors (CSFs) of the organization was also used by researchers studying warehouse performance through developing a process performance model (Chen et al., 2017).

All in all, it was observed that the overall warehouse performance (OWP) was studied using different methods, most commonly DEA and its variations were used. This method compares the outputs produced with the inputs of a process. However, using DEA does not eliminate the fact that some KPIs are dependent on each other. Hence the factors making up the overall score of the warehouse are not stand-alone independent factors. Furthermore, because each DMU is affected by different exogenous effects, it is impossible to ensure that the most efficient DMUs can be compared to the least efficient ones (Pinto et al., 2017). Not only that, but the focus is mainly on outbound process KPIs as was observed in the literature and little research regarded inbound KPIs.

This is problematic as the shortcomings of the outbound KPIs may be caused by poor inbound results since they are dependent on each other.

To overcome these issues, an overall metric must consider key KPIs that match the CSFs and at the same time have factors that are independent of each other to give a more accurate estimate of the OWP. One metric that has been widely accepted as an overall effectiveness metric in the manufacturing industry is called overall equipment effectiveness or OEE.

2.2 OEE in Literature

2.2.1 OEE Background

The first time OEE was introduced was in 1988 by Nakajima as a part of total production maintenance (TPM). OEE was developed as an overall metric that measures effectiveness in equipment. TPM focuses on production equipment, especially for highly automated processes in production. Ever since its development, it has been used by various manufacturers, each slightly adapting it for their industry. OEE defines the major losses that reduce production effectiveness. This helps in suggesting areas of improvement and directing optimization to the losses that have the biggest impact on OEE (Muchiri & Pintelon, 2008). OEE can also indicate if improvements were fruitful for the overall performance, analyzing the increase or decrease of OEE after the implementation of certain improvements can give an idea if that change was impactful over time (Dal et al., 2000b).

There is some debate about whether OEE is a measure of effectiveness as the name implies or an efficiency metric. Effectiveness is described as a process feature that shows the degree to which the process output complies to the standards. It shows whether something has been done correctly. Efficiency, on the other hand, is a process characteristic that indicates how well a process generates the desired result while using the least number of resources. It shows how well the process was done in terms of the resources used (Muchiri & Pintelon, 2008). OEE can be described as a measure of total equipment performance, or the extent to which the equipment is performing as intended (Williamson, 2006). The OEE tool is used to identify losses that impair the efficiency of equipment. These are activities that consume resources but do not generate value. Manufacturing disruptions, according to (Jonsson & Lesshammar, 1999)cause the losses. They divide disturbances into two main categories. Chronic disruptions are subtle and difficult to detect, and they can be caused by a variety of factors. Sporadic disturbances, on the other hand, are more noticeable since they occur frequently and have significant departures from the normal condition.

The metric Is composed of three elements independent of each other that together make up the overall metric. These are *Availability, Performance,* and *Quality.* The three components measure distinctly different things. Where availability is defined as the ratio of time the machine was actually used to the total time it was planned to be in use after deducting the scheduled downtimes. Performance describes how fast the machine is running, this is calculated as the ratio between the actual production to the standard or nominal production during the actual time the machine was working. Finally, quality measures the percentage of conforming units produced (García-Arca et al., 2018).

The six big losses described by Nakajima can be summarized as follows:

Losses that have an impact on *Availability* include:

- Equipment breakdown or failure, which results in time and quality losses as well as major shutdowns.
- Setup and adjustment delays, these are brought on by the shift in production needs.

Losses that have an impact on *Performance* include:

- Idling and minor stoppage losses, this is when a machine is idle, or output is temporarily interrupted by a problem.
- Reduced speed losses, this is the discrepancy between the equipment's design speed and its actual operating speed.

Losses that have an impact on *Quality* include:

- Decreased yield from the moment of startup to stabilization
- Losses from reworks and quality flaws

To calculate OEE, the following formulas are used:

$$OEE \ = \ Availability \ x \ Performance \ x \ Quality$$

Where:

$$\text{Availability} = \frac{\text{Planned production time} - \text{ Unplanned downtime}}{\text{Planned production time}}$$

$$\text{Performance} = \frac{\text{Actual Amount of production}}{\text{Planned amount of production}}$$

$$\text{Quality} = \frac{\text{Actual amount of production} - \text{ Non} - \text{accepted amount}}{\text{Actual amount of production}}$$

These definitions are based on the work of (de Groote, 1995). Other researchers may calculate *Availability* based on the total available time as was done by Ivancic in 1998 when he developed a new metric named Total Equipment Effectiveness Performance (TEEP) (Muchiri & Pintelon, 2008). That was done to study the impact of preventive maintenance on the overall performance of the equipment. Figure 5 depicts a general framework for OEE from the works of (Pintelon et al., 2000) as it shows how the total time being effectively used is arrived at after the six major losses.

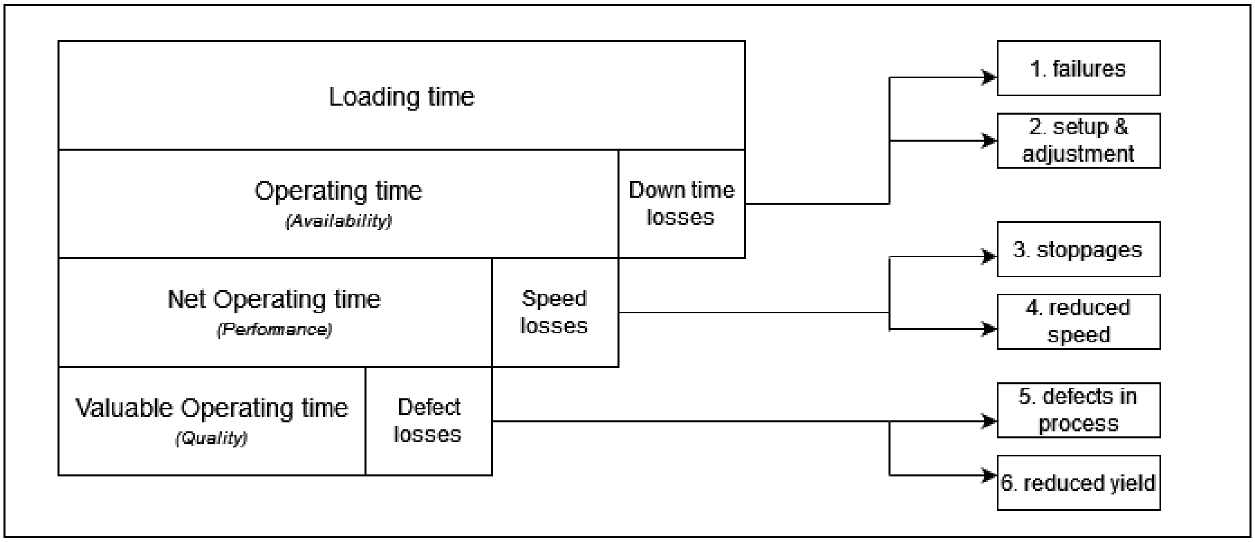

Figure 5 General OEE framework (TPM view); page 26 (Pintelon et al., 2000)

The general framework of OEE shown above may not be applicable to all industries and all equipment. It is also worth noting that the traditional OEE model made by Nakajima is meant to measure the effectiveness of one machine rather than an entire factory for example. This led to the evolution of OEE.

2.2.2 OEE Evolution

OEE has several industrial applications in different sectors of the economy today. Although the original OEE concept served as the foundation for gauging effectiveness, manufacturers modified OEE to meet their unique industrial needs. Additionally, in literature, the term OEE has been changed to other terms that refer to the idea of application. OEE has expanded to include total equipment effectiveness performance (TEEP), overall factory effectiveness (OFE), overall plant effectiveness (OPE), overall throughput effectiveness (OTE), production equipment effectiveness (PEE), and overall asset effectiveness (OAE) (Muchiri & Pintelon, 2008).

The general definition of OEE and the major losses must be broken down to translate the input data from the monitoring of the processes into an OEE valid for the line under study. Naturally, different industrial areas have different needs. This led to practitioners and researchers modifying the original OEE to suit their areas.

2.2.2.1 Adding Weights

One of the main shortcomings of traditional OEE is that it treats each of the three factors equally. Hence, the availability, performance, and quality of the unit under study all have the same weight in calculating the overall effectiveness. While this may be suitable for some manufacturers, others may have different agenda. A pharmaceutical manufacturer for example may put a higher value for their products to be conforming to the standards they should be. For this example, the quality factor would be influential on the overall score of the unit under study. To overcome this issue, a modification to the traditional OEE was made by adding weights to each factor.

Production Equipment Effectiveness (PEE) which was developed by Raouf in 1994, adds weights to the calculation of OEE. Additionally, it distinguishes between discrete-type production activities and continuous process operations. The distribution of weights among the various components of overall effectiveness is the primary distinction with OEE. Contrary to the fundamental premise of OEE that the three criteria have the same weight, it assumes that quality has a distinct weight from performance and availability for example. This can be evident in the pharmaceutical industry as an example.

2.2.2.2 Overall Factory Effectiveness

Another matter of concern regarding traditional OEE is that it is meant for one equipment rather than the entire production line. The Overall Factory Effectiveness (OFE) measure was developed to measure effectiveness of an entire factory that involves several production processes and stations. The objectives of OFE include optimizing the sequence of orders, works, or jobs; ensuring a balanced line and smooth flow of work by integrating micro-scheduling with overall plant scheduling; and synchronizing the production schedule with planned setup time, qualification time, and downtime through a tighter connection to enterprise planning systems and infinite capacity schedule (Williamson, 2006).

For the OFE measurements, there has never been a standard methodology. According to (Scott & Pisa, 1998), there cannot be a single indicator because different plants have various objectives. They recommend developing composite metrics that ask for key objectives, such as cycle-time efficiency, on-time delivery, capacity utilization, rework rate, yield rate, etc. The key to achieving company objectives is defining the success measuring criteria and applying the desired weighting variables. An overall result is generated for these metrics (Muchiri & Pintelon, 2008).

2.2.2.3 Identifying relevant losses

Another point of criticism about OEE is that the losses identified by Nakajima may not be sufficient or suiting for every industry. For this reason, researchers and practitioners have modified the general framework to include other losses that impact OEE. These include lack of demand, detailed operational losses, external affecting factors like the delay of supplies, and internal factors like the starving of machines. While there is no agreement between industries on what should be included in calculating a modified OEE, there is certainly an agreement on categorizing the losses in a way that identifies the major losses causing a sub-optimal performance. Modified OEE models also included other factors than availability, performance, and quality as will be shown in the next section.

2.2.2.4 OEE in Different Areas

Naturally, as OEE evolved in the literature and its application has been modified in practice, the scope of OEE has widened to include other areas than manufacturing. As mentioned earlier, these adaptations were made to accommodate to the needs of different industries and adapt the traditional OEE from the criticism. Nowadays, OEE is considered a must-have for firms maintaining an Industry 4.0 standard according to Walker Reynolds, president and solutions architect at 4.0 Solutions in the United States.

Ng et al. (2020) performed an extensive systematic literature review of OEE approaches. To our knowledge, this is the most comprehensive overview of the current literature in OEE as 186 articles were used for the review. In their paper, the academic interest evolution was explored and a list of existing OEE models was made as well as areas where OEE is an emerging topic. The results show that OEE was modified over time by adapting it to other industries or by including other variables in the calculation. Table 3, adapted from the extensive table the authors made, shows the development of OEE models and adaptation to different fields.

Source	Model Name	Brief Description
(Huang et al., 2002; Muthiah & Huang, 2007)	Overall Throughput Effectiveness	Calculates the productivity of a manufacturing system; measures the factory level performance; identifies the bottleneck and hidden capacity
(Nachiappan & Anantharaman, 2006)	Overall Line Effectiveness	Measures the productivity of a line manufacturing system.
(Domingo & Aguado, 2015)	Overall environmental equipment effectiveness	Identifies losses due to sustainability, based on the calculated environmental impact of the workstation
(Puvanasvaran et al., 2017)	Modified OEE	Includes losses associated with human factors and usability (the frequency of setup and changeover process)
(Nakhla, 2018)	Extended overall equipment effectiveness	Evaluates the entire process considering human resources and equipment Performance. It is applied in medicals activities of operating rooms.
(García-Arca et al., 2018)	OEE to transport management	Improves efficiency in road transport by adapting OEE to transport management.
(Larrañaga Lesaca et al., 2017)	Strategic equipment effectiveness Operational equipment effectiveness	A global measure of the effectiveness of an integrated electrical system.
(Pinto et al., 2017)	OEE of port terminal	Identifies the most efficient terminal, addressing either manageable or unmanageable factors

Table 3 List of OEE based models from (Ng Corrales et al., 2020)

It is noted after analyzing different OEE approaches that the traditional formula is adaptable to various areas to measure effectiveness in general rather than only in production equipment. To create a modified OEE, an extensive study of the process under examination and the factors to be included in calculating OEE for that domain must be made. As can be seen in Table 3, OEE expanded to include whole production lines. Not only that, but additional factors have been added such as sustainability to identify the environmental effect as was done by Domingo & Aguado (2015).

Puvanasvaran et al. (2017) explore the human factor in OEE calculation as they propose a modified OEE to include the impact of the frequency and efficiency of changeovers from a human factors perspective. Nakhla (2018) also includes the human factor in her extended OEE in the medical field for operating rooms.

Other areas where OEE was applied includes road transport, integrated electric systems, and port terminals. These are just some examples of the diverse areas where OEE was applied in after modifications. Hence it is safe to say that the concept is generally acceptable and is used in practice not only in manufacturing.

Future research on OEE can be applied to the logistics industry and used to create environmental variables like the carbon footprint created during a particular process. OEE can be used in supply chains to evaluate a warehouse's productivity when it comes to moving cargo around. When it comes to the availability, effectiveness, and quality of the services received, OEE can be utilized in the service industry to measure customer satisfaction. To further illustrate a company's overall productivity, an OEE-based model can be added to a balanced scorecard. All of these actions give a broad view of the company's operations and accomplish the two key production goals of r'ising productivity and reducing waste (Ng Corrales et al., 2020).

2.2.2.5 OEE in Logistics

The use of OEE for logistical processes has been explored by several authors where an adaptation to the traditional OEE was made. However, this does not come close to the number of publications of OEE usage in manufacturing. Examples of areas where OEE was used in logistics include road transport, unloading of inbound trucks and urban freight transportation systems. These examples are contributions where the authors were pioneers in using an adapted OEE for their respectful field of study.

When observing the use of OEE in logistics, it is noted that identifying and classifying the losses of the specific case study and aligning it with the six big losses marks the initial phase of adapting OEE. It is worth noting that the term "logistical uptime" (as will be used in this report) can be defined as a measure of reliability or efficiency that expresses either the duration of uninterrupted system availability in time units, or the percentage of total time for which there was uninterrupted system availability (Rogers, 2021). Hence the use of OEE in logistics can be considered a measure of logistical uptime.

Some authors merely define the availability, performance, and quality variables from their case study by adapting the general definitions to specific KPIs and variables in that area. This was evident in the adaptation of OEE in road transport as a list of KPIs based on literature search and relevance to the logistical success was made. These were integrated when measuring the effectiveness of road transport systems by using the adapted OEE (García-Arca et al., 2018).

Other authors included an additional variable to OEE. In their paper, Ng Corrales et al. (2022) added the variable of punctuality to make a new metric, Overall Process Effectiveness (OPE). The punctuality of the truck arrival to get unloaded was an external factor that affected the logistic uptime of the overall system. The inclusion of the new punctuality factor in calculating the effectiveness of the logistical process showed that the real effectiveness was different to what the traditional OEE would reveal.

In a case study of applying OEE to freight transport, the application of OEE to measure the effectiveness of urban freight transport systems provided the best input for optimization of their systems by using a multi-objective mathematical model. Comparing the models results of OEE, classic, and profitability optimizations of the model showed that the best results were obtained from the OEE measurement. Not only that, but the authors noted that the quality indicator optimization guarantees a good overall performance of freight transport (Muñoz-Villamizar et al., 2018).

Chapter 3: Methodology

3.1 Research Design

In the logistics and supply chain management (SCM) field, practitioners face new challenges daily due to the growing demand of highly efficient supply chains. To deliver the promises made by organizations, tight margins often need to be met. To deliver a product to a customer withing 24 hours of ordering, all the hurdles that may occur in the process form problems to be solved by practitioners. These growing challenges are addressed by firms using theoretical models and emerging technologies such as the techniques currently used in the Industry 4.0 approach. To accommodate these issues and to facilitate the problem solving, knowledge is vital for innovative solutions. This is of importance to the academic field as there is a gap between the academic research and practitioners' experience in the SCM field (Sandberg et al., 2022). In practice, supply chain experts seldom rely on pure academic research when dealing with the challenges faced.

The problem faced by the company of the case study lies in the lack of an overall metric to assess their effectiveness in operating the warehouse. However, the objective of this research is not limited to providing an overall metric for a specific warehouse. This paper explores the use of a new metric in the warehouse industry. A primary objective of this research is to contribute to the academic field of OEE evolution. Providing a ground for further research into the use of OEE in warehousing should also be achieved. Identifying the practicality and usefulness of this metric for warehouses is to be evaluated.

Choosing a methodology approach that align with these goals and satisfies research objectives is vital. The following subsections motivate and further describe the methodology chosen for this paper.

3.1.1 DSR vs AR

The most prominent collaborative approach used in the supply chain context is action research (AR). However, it has been criticized for its focus on solving practitioners' problems and the lack of its application to academic research and theory (Shani & Coghlan, 2019). A similar approach that aims to solve real-life problems through research is Design Science Research (DSR). In this domain, the goals of the two research methodologies are similar in that they both aim to solve problems and advance knowledge. However, each method's rules adhere to particular flows. When conducting action research, the researcher considers the circumstances and goals of the study, gathers data, solicits feedback, analyzes the data, develops action plans, carries them out, assesses them, and then delivers the findings. On the other hand, the researcher using design science research identifies a practical problem, designs and evaluates an artefact, on occasion implements and assesses the artefact's range of application and potential for generalization to a class of problems, identifies and analyzes the artefact's theoretical and practical contributions, and finally presents the findings.(Collatto et al., 2018). Hence it is argued that DSR serves less as a "consultant" role and in turn contributes more to academic research and the knowledge base.

Another key distinction between DSR and AR is that DSR, unlike AR, does not assume a specific client nor a partnership between researchers and the client. But at the same time the generated artefact often seeks to handle a class of problems in a way that is helpful in addressing particular problems of a certain client. Therefore, it is possible to contend that DSR has potential clients who are "The set of all members of the generalized class of all people or organizations who could potentially be motivated to solve instances of the generalized class of problem(s)" (Venable, 2009) addressed by the DSR outcome or artefact. (Iivari & Venable, 2009)

Table 4 shows how Dresch et al. (2015) compare DSR to AR from various aspects. This gives further insight to decide the right methodology for this project.

Characteristics	Design science research	Action research
Objectives	Develop artefacts that enable satisfactory solutions to practical problems	Solve or explain problems of a given system by generating practical and theoretical knowledge
	Design and recommend	Explore, describe, explain, and predict
Main activities	Define the problem Suggest Develop Evaluate Conclude	Plan actions Collect data Analyze data and plan actions Implement actions Evaluate results Monitor (continuous)
Results	Artefacts (constructs, models, methods instantiations) and improvement of theories	Constructs Hypothesis Descriptions Explanations Actions
Type of knowledge	How things should be	How things are or how they behave
Researcher's role	Builder and/or evaluator of the artefact	Multiple, due to the action research type
Empirical basis	Not mandatory	Mandatory
Researcher-researched collaboration	Not mandatory	Mandatory
Implementation	Not mandatory	Mandatory
Evaluation of results	Applications Simulations Experiments	Comparison against the theory
Approach	Qualitative and/or quantitative	Qualitative
Specificity	Generalizable to a certain class of problems	Specific situation

Table 4 DSR vs AR; page 95-96 (Dresch et al., 2015)

The research of using OEE as a performance measure for warehouses is a new area, hence it is not expected that an immediate implementation of the artefact to solve the problem will take place. Moreover, the case study for this project involves the inbound part of a warehouse, however, the aim is to develop an OWP measure through OEE. Hence it is crucial for the artefact to be generalizable to other warehouse areas in order to arrive at a complete OWP model. Moreover, a recommendation to the warehouse industry of using or not using an OEE based model for performance measurement is expected of this project. This makes DSR a more suitable methodology for such research. Additionally, due to the lack of papers addressing the use of OEE in warehousing, comparing results to theory is challenging. Action research calls for an evaluation of results by comparing it with existing theory, whereas DSR evaluates the outcomes in various ways that are possible in this study as will be discussed further.

It is worth noting that design science research artefacts are rarely fully developed information systems that are put to use in real-world situations. Instead, artefacts are innovations that specify the theories, methods, tools, and products that make it possible to successfully and efficiently analyze, develop, execute, and use information systems (Denning, 1997) as cited by (Hevner et al., 2004). Since this research area is young it makes sense for the work to move more towards this direction as a start for the research of applying modified OEE models in warehousing.

For these reasons, a DSR approach will be used for this thesis as it is argued to match the aim of this paper, providing an OWP framework as an innovative artefact using a modified OEE not just for one client, but for any fitting warehouse similar to the case study example.

3.1.2 DSR Definition and Objectives

Design science is defined as “the scientific study and creation of artefacts as they are developed and used by people with the goal of solving practical problems of general interest” (Johannesson & Perjons, 2014). Hence Design Science Research (DSR) is a paradigm for problem-solving that aims to advance human knowledge through producing original artefacts. Simply said, DSR aims to advance the knowledge bases of technology and science by developing novel artefacts that address issues and improve the environments in which they are implemented (Hevner et al., 2004). The outcomes of DSR comprise both the newly created artefacts and design knowledge (DK), which offers a deeper understanding of how and why the artefacts improve (or disrupt) the pertinent application contexts through design theories (Brocke et al., 2020).

A DSR research project's objective is to raise the bar of what people and organizations are capable of by creating brand-new, inventive artefacts that are represented through constructs, models, techniques, and instantiations (Gregor & Hevner, 2013). An object created by humans with the goal that it be used to solve a practical issue is referred to as an artefact in this definition. Some artefacts are tangible things, like hammers, automobiles, and hip replacements. Other artefacts come in the form of sketches or blueprints, such as a building's design by an architect. A method for designing databases is an example of a method that might be considered an artefact. All of these objects have one thing in common: they assist users when they run into difficulties with a particular exercise (Johannesson & Perjons, 2014).

3.1.3 Implementing DSR

DSR researches relevant issues in the real world across a range of application disciplines. According to research, there is a demand for solutions to be empirically tested with individuals working in businesses that employ particular technologies. A DSR project's starting point is often the examination of the business environment and the identification of specific demands that must be met. However, there are also circumstances where needs have previously been examined and can be inferred from current research such as further research recommendations. In order to determine how much design knowledge is already accessible to address a certain challenge, DSR analyzes the academic knowledge base. These can be theories, frameworks, tools, or design objects like constructions, models, techniques, or instantiations. If there is information already accessible to address an issue, and can be merely applied by replication, then this does not fall under DSR. To clarify the difference, DSR seeks to develop an original solution to the issue, which typically draws on already-existing components of a solution and merges, revises, and expands existing design expertise. Hence the artefact developed is a combination of existing theories and technologies for example, not just a reproduction of a previously used technique, but more of a construction from various knowledge sources. The "build" and "evaluate" activities that make up the design activities often come after several iterations. During a DSR study, a variety of research techniques are used, including those that are well known in social science research, like focus groups, surveys, interviews, and literature reviews. (Brocke et al., 2020)

The methods used for this research are essentially literature review, observations of a case study when studying the problem and interviews to evaluate the created artefact. More detail about the methods used for each stage will be provided throughout the coming sections.

3.1.3.1 Design Science Research Methodology

The process model put forth by Peffers et al. (2007) is the one that is most frequently cited in the literature (Brocke et al., 2020). Figure 6 depicts the design science research methodology (DSRM) process paradigm. It consists of six steps: problem identification and motivation, definition of solution objectives, design and development, demonstration, assessment, and communication.

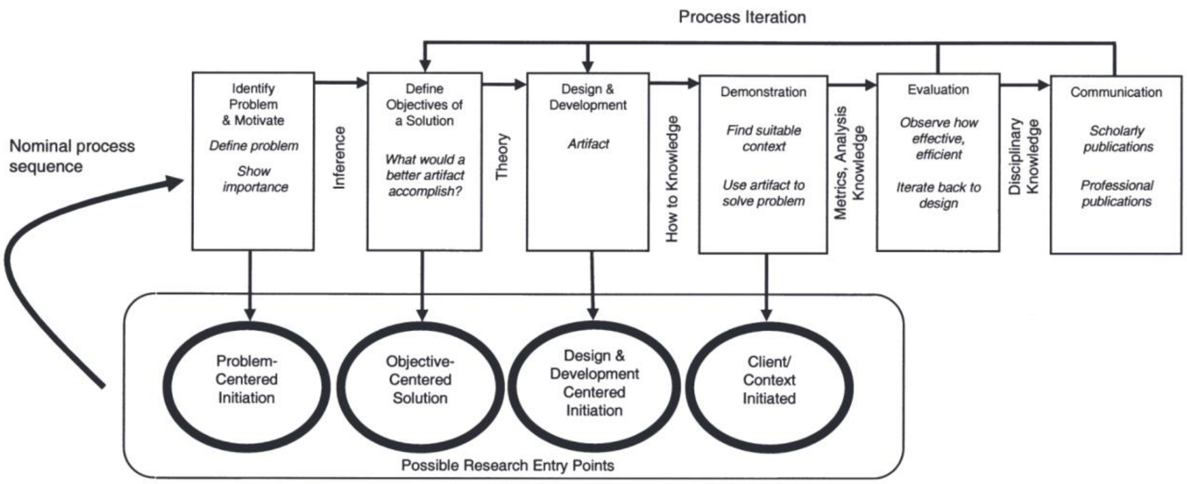

Figure 6 DSRM process framework, Source: (Peffers et al., 2007)

These six steps are briefly described here using the works of Peffers et al. (2007):

1) Problem identification and motivation: Defines the exact research issue and defends the necessity of a solution. It may be helpful to conceptually break down the problem so that the solution may adequately reflect its complexity since the problem description will be utilized to construct an artefact that can effectively give a solution. In addition to inspiring the researcher and the research's audience to seek the solution and accept the findings, justifying the worth of a solution also aids in comprehending the thinking behind the researcher's grasp of the problem. Knowledge of the problem's current status and the significance of finding a solution is one of the resources needed for this task.

2) Definition of solution objectives: Uses the problem definition and the understanding of what is feasible and possible to arrive at the goals of a solution. The goals can be qualitative, such as a description of how a new artefact is intended to suggest answers to problems not previously addressed, or quantitative, such as terms in which a desirable solution would be better than present ones. The goals should be logically deduced from the problem description. The state of the problem, any existing solutions, and the effectiveness of those solutions are resources that are needed for this.

3) Design and development of the artefact: Here the artefact is made. These artefacts could be constructions, models, procedures, or instantiations. A design research artefact can theoretically be any created object in which a research contribution is incorporated into the design. This activity entails deciding on the architecture and expected usefulness of the artefact before building it. Knowledge of theory that can be used to a solution is one of the resources needed for transitioning from objectives to design and development.

4) Demonstration: This stage shows how the artefact can be used to resolve one or more instances of the problem. This could entail applying it to an experiment, simulation, case study, or other applicable task. Effective understanding of how to use the artefact to solve the problem is one of the resources needed for the demonstration.

5) Evaluation: Observing and evaluating how well the artefact contributes to a solve the problem takes place here. Comparing a solution's goals to the results that were really obtained when the artefact was used in the demonstration is the focus of this activity. Evaluation can take many different shapes depending on the artefact and the problem area. It could consist of things like a comparison of the functionality of the artefact with the solution goals from activity 2, objective quantitative performance measurements like budgets or output, the outcomes of customer satisfaction surveys, simulations, or budgets or output. It can contain measurable indicators of system performance, like availability or reaction time. Conceptually, such an assessment may include any pertinent logical or empirical support. At the end of this activity, the researcher can choose to go on to communication and leave further improvement to following projects, or they can go back and try to increase the effectiveness of the artefact in activity 3. Whether or not such iteration is possible may depend on the nature of the study environment and the time available.

6) Communication: It involves informing researchers and other pertinent audiences, such as working professionals, about the problem and its significance, the artefact, its uniqueness and utility, the rigor of its design, and its effectiveness. Similar to how the nominal structure of an empirical research process (problem definition, literature review, hypothesis development, data collection, analysis, results, discussion, and conclusion) is a typical structure for empirical research papers, researchers may use the structure of this process to structure the paper in scholarly research publications. An understanding of the discipline culture is necessary for communication.

DSR research follows several guidelines that define it as a DSR paper and differentiate from a mere display of technical skills. These guidelines presented by Hevner et al. are shown in Table 5. Throughout this thesis, these guidelines will drive the work synthesized and presented. This includes the beginning of the problem definition and continues throughout the artefact development, evaluation, and communication of the DSR results.

Guideline	**Description**
Guideline 1: Design as an Artefact	Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.

Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact.
Guideline 6: Design as a Search Process	The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Table 5 DSR guidelines, Source: (Hevner et al., 2004)

3.1.4 DSR Evaluation

One of the guidelines that is worth expanding upon is evaluation. DSR gains an advantage over AR by having a “build and evaluate” approach where the researcher learns throughout the process and does not only have the aim of serving a particular client. Several evaluation methods are proposed by Hevner et al. in Table 6 where a researcher is not necessarily bound to them, however it resembles a starting point and gives an idea of how artefacts can be evaluated.

Design Evaluation Methods	
1. Observational	Case Study: Study artefact in depth in business environment
	Field Study: Monitor use of artefact in multiple projects
2. Analytical	Static Analysis: Examine structure of artefact for static qualities (e.g., complexity)
	Architecture Analysis: Study fit of artefact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artefact or provide optimality bounds on artefact behavior
	Dynamic Analysis: Study artefact in use for dynamic qualities (e.g., performance)
3. Experimental	Controlled Experiment: Study artefact in controlled environment for qualities (e.g., usability)
	Simulation — Execute artefact with artificial data

4. Testing	Functional (Black Box) Testing: Execute artefact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artefact implementation
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artefact's utility
	Scenarios: Construct detailed scenarios around the artefact to demonstrate its utility

Table 6 DSR evaluation methods, Source: (Hevner et al., 2004)

For this research an analytical and descriptive evaluation takes place. The developed artefact is presented to experts in the field from a case study and the usefulness and utility of it is examined. This is done using semi-structured interviews to present relevant results and give the experts more room to evaluate based on their grounds and needs which align with the DSR approach, solving a real-life problem. The reason for choosing this method of evaluation is because of the absence of research using OEE in warehouses. Hence this paper marks the very beginning of the exploration of OEE in warehouses. Therefore, it is more sensible to first explore its utility through the input of experts in this field before further diving into more complexity that may not be needed if the initial response was negative towards the use of OEE in warehouses.

3.1.5 Design Knowledge Contribution

There are different contribution types based on the knowledge maturity and abstractness of the artefact developed. Table 7 shows the different types of knowledge contributions from DSR as described by Gregor & Hevner (2013). It is important to know in which direction the research is going. The goals to be achieved by the researcher can help steer into a certain knowledge contribution type. For this research, since it is of explorative nature and aims to present a novel idea to be further developed in future works of academics and practitioners, it is argued that a Level 2 contribution is more consistent with the goals of this paper.

	Contribution Types	Example Artifacts
More abstract, complete, and mature knowledge	Level 3. Well-developed design theory about embedded phenomena	Design theories (mid-range and grand theories)
	Level 2. Nascent design theory—knowledge as operational principles/architecture	Constructs, methods, models, design principles, technological rules.
More specific, limited, and less mature knowledge	Level 1. Situated implementation of artifact	Instantiations (software products or implemented processes)

Table 7 DSR knowledge contribution types, Source: (Gregor & Hevner, 2013)

They also introduced a framework for DSR knowledge contribution as shown in Figure 7.

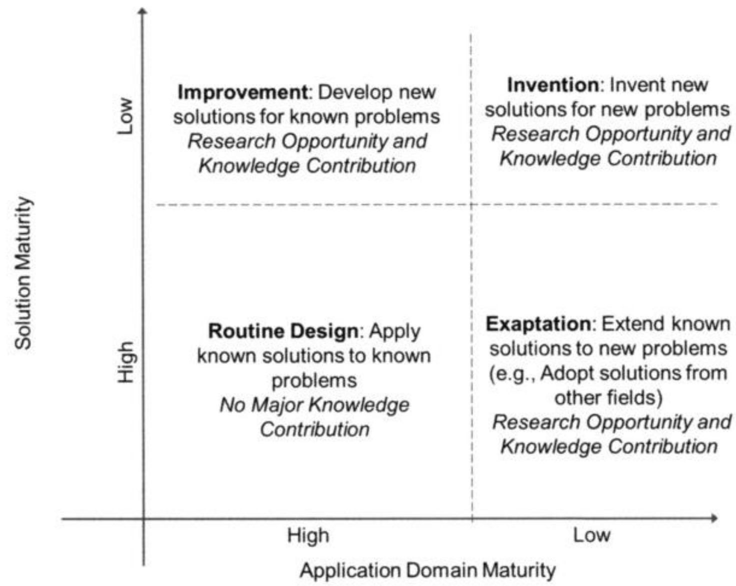

Figure 7 DSR knowledge contribution framework, Source: (Gregor & Hevner, 2013)

There may be research situations when the artefacts needed in a field are either unavailable or not at their best. Effective artefacts might still exist in related problem domains, though they might need to be "exapted" to the brand-new problem context. An exaptation of a solution means that the use of the solution is different to the domain that it was intended for. Contributions in this quadrant expand upon or improve upon existing design knowledge in one discipline so that it may be used to a new application area (Gregor & Hevner, 2013).

Following this framework and the description above, it is clear that adapting OEE to come up with an OWP measure is an example of an exaptation. Hence it is expected that at the end of this research, greater research potential is identified and a knowledge contribution of the use of OEE in warehousing is achieved.

3.2 Data Collection and Analysis

As previously mentioned, a variety of research techniques are used in DSR, these include interviews and recording observations amongst other well-known methods. The following subsections describe how the research took place and shed light on the methods used to accomplish a DSR project.

3.2.1 Unstructured and Semi-structured Interviews

After conducting a literature review, the ideas of interest become clearer, and units of analysis arise. Since a DSR approach is being used to solve a real-life problem, there must be a connection between the literature and the world of practice. After identifying the different warehouse KPIs and the variety of warehouse types and their functions, questions arise regarding what is to be included in the OWP artefact and why that is important and how it can be utilized. To answer these questions, a mixture of unstructured and semi-structured interviews take place. Two types of semi-structured interviews were made, the first set of interviews aimed to understand the system and collect the relevant KPIs and CSFs. The second set of interviews had the purpose of evaluating the artefact. In addition to those interviews, input data for the model is collected to test the artefact quantitively as will be discussed.

3.2.1.1 Data Collection to Build the Artefact

Firstly, before all the goals of research are in place and clear to the researcher, unstructured interviews with employees at the case study company provide an insight on the composition of the warehouse and the different goals of the stakeholders. According to Zhang & Wildemuth (2009), unstructured interviews are particularly helpful when one wishes to study a specific phenomenon in a particular cultural context in depth. Additionally, they are best used when doing research using an interpretive paradigm, which makes the assumption that participants in the study's environment socially construct reality. On the basis of this underlying premise, one should try to comprehend the phenomenon of interest from the viewpoints of the many parties involved. Appendix A outlines questions that start these interviews where the general purpose is to grasp the various elements of the system.

After that, and upon a deeper understanding of the case study starting to develop, semi-structured interviews take place to pinpoint the relevant issues that should be considered when building the artefact. Semi-structured interviews are used in the current study to enable interviewees to actively share their knowledge. Although the interviews are guided by a protocol with formulated questions, additional questions might be asked to refine the participant's initial response (Karlsson, 2016; Zowghi & Coulin, 2005).

Mainly, the goal of the interview is to provide a relevant artefact, understand the problems occurring and using these statements to build a problem-solving artefact. As mentioned in the previous chapter, the case study happens within a logistics service providing company. The goals associated with their activities and the problems specific to their operations are explored in these interviews.

The interview protocol is shown in Appendix B where the semi-structured interview is divided into four sections. The participant's general information, such as work title or prior relevant experience, is gathered in the first section as well as an indication if they are aware of OWP. The second section asks broad questions concerning the factors of concern when evaluating a service provider's performance. These questions revolve around KPIs and their importance from a service provider’s perspective, not only that but also questions about certain measures and how they are recorded are asked. However, the third section asks the participant about specific OWP and OEE related questions. The purpose of these questions is to further understand and identify CSFs from a service provider’s point of view and to identify the importance of having an overall measure and the possible utility of this number. The final section includes closing questions that provide the participant a chance to address any problems or further discuss certain ideas.

3.2.1.2 Data Collection for Evaluating the Artefact

The qualitative evaluation of the artefact is done using semi-structured interviews to get a more elaborate evaluation and allow for recommendations and suggestions for further research. It is argued that assessing the artefact qualitatively relying on expert views should precede any quantitative evaluation for a similar research in a young field. This is due to the exploratory nature of this study and the lack of research of OEE in warehouses. Appendix C shows the template used to carry out these interviews.

As for the quantitative evaluation, data was collected to fill in the model created in Excel. A daily breakdown of the warehouse performance and KPI forms was used to test the artefact and obtain calculations that will be used to analyze and evaluate the performance of OEE as an OWP measure. Appendix D shows the data collected for this purpose.

3.2.2 Informal Methods

In addition to the interviews, informal methods outlined a significant part of familiarizing with the case study. Numerous detailed warehouse tours provided a better overall understanding and visualization of the warehouse. Watching workers in the field firsthand gives a more in-depth grasp of the warehousing operations. Most importantly, observing the three different partners in the field and asking about the interactions between them helps the researcher distinguish between their goals and identify CSFs for the specific artefact under development. This provides increased relevance for a service provider artefact rather than a more general and less accurate one.

Other activities include attending KPI reporting meetings where the performance of the company is discussed on a daily basis. This helps identify the issues of concern even more and gives room to ask questions to employees of different functions to clarify some areas that seem to be vague. Lastly, small side talks and quick questions to employees in the organization helped to allow the researcher to live the full experience and be connected with the various aspects of the company which helps increase the relevance of the artefact.

Data that will be used in the artefact later on merely for presentation purposes is taken from company records from their data base. These are pointed out in the following chapter.

3.2.3 Data Analysis

The data collected is used for building the artefact as well as evaluating it. The results of the first set of interviews and informal methods of collecting notes during warehouse notes are analyzed using a content analysis approach. For the qualitative analysis of the interviews, various aspects of the evaluation criteria are evaluated through assessing the interview responses and identifying if any particular points are stressed by the participants. A narrative analysis of the participant responses indicated how they view the utility of the artefact, its ease of use, and relevance.

Quantitative analysis of the model includes comparing OEE values calculated with dock-to-stock times score to test if there is a correlation between the OEE values calculated and inbound performance. JASP software is used to test for correlation. Moreover, a sensitivity analysis of the artefact developed is done using Excel as will be discussed in the evaluation chapter.

3.3 Research Validity

In case study research, it is critical to demonstrate validity and reliability (Karlsson, 2016). The corresponding dimensions are briefly discussed in the following paragraphs.

The validity of the construct is demonstrated by information gathered from various sources. Data triangulation entails interviewing personnel from various organizational divisions, assessing demands, and identifying new requirements from various functional perspectives. Table 8 shows an overview of the participants for the interviews.

Expert interviews are used to validate the specified criteria and, ultimately, the preliminary OWP and OEE models for internal validity. External validity will result from assuming that industry standards can be developed and applied to other logistics service providers based on research with an industry leader.

Furthermore, the basic OEE model is modified based on additional domain-specific requirements identified during the interviews to ensure applicability. Finally, interview methodology that directs interviews toward a specific study goal ensures the reliability of the research (Karlsson, 2016).

Participant	Title	Role in the Company
Participant A	Integration Manager	Integrating newly introduced software into the system and ensuring operational activities go live with the new software and changes.
Participant B	Control Room Manager	Oversees all control room specialists and overviews the WMS related errors and downtime recordings
Participant C	Business Analytics Manager	Developing comprehensive tools and strategies that allow raw data to be transformed into business insights

Table 8 Interviewees overview

To validate the model quantitatively, a sensitivity analysis ensures that the changes of input result in a significant change in the output of the model. Moreover, in the last stages of the research and upon the qualitative evaluation, a test of correlation between OEE and dock-to-stock scores takes place to ensure the validity of OEE as an OWP measure.

Chapter 4: Artefact Development

The first thing to consider when developing this artefact is answering the question: “What is it measuring?” As discussed earlier, there are many warehouse KPIs that might be overwhelming for some. Many KPIs are inter-related and often one affects the other. Not only that, but there is always a specific set of KPIs that are more critical than others. These usually guarantee that the organizational strategy is being implemented and tracked.

This study was done on the inbound part of the warehouse, hence narrowing down the scope of selected KPIs. Knowing what KPIs are relevant for inbound and what is prioritized by firms of similar nature to CEVA marks the beginning of developing an overall metric. CEVA, which is a logistics service provider, operates a wide range of warehouses. The warehouse under study belongs to a large online retailer. For such warehouses, it is of great importance for the products to be swiftly available for any customer order to avoid delivery delays.

It is important to note that the work done in this research is most relevant to firms that have similar products as the case company. The warehouse owner, which is an online fashion retailer, deals with mainly non-fragile products that do not have expiration dates like foods for example. Hence this makes the KPIs and CSFs of importance in this study less relevant for a fresh produce distributor for example. Therefore, it is worth differentiating between the CSFs and KPIs of different warehouse types.

The artefact development starts with combining the knowledge gained from the literature review and the interviews that took place. The interviews and field visits resulted in a deeper understanding of the case under study and were used to identify the relevant KPIs and CSFs to be incorporated into the artefact. A description of these interviews was previously discussed and further discussion about the utility of the artefact will follow in chapter 6.

4.1 Identifying OWP Factors

To begin creating the artefact, the factors to be included in the modified OEE model need to be identified. To do this, relevant warehouse KPIs and the CSFs of service providers are explored to then be combined into one overall measure.

4.1.1 Stakeholders in a Warehouse

As mentioned earlier, the perspective of this OWP measure is from a service providing firm. The warehouse under study has three main stakeholders that compose the overall system. Firstly, the warehouse owner, which is the online retailer in this case study. This stakeholder designs the warehouse and provides the warehouse management system (WMS). CEVA, which is the second stakeholder here, operates the warehouse. Hence, the majority of the workers in the warehouse are CEVA operators. Thirdly, the mechanization partner is responsible for building the warehouse components and taking care of the mechanical parts and maintenance. This split of responsibilities is important to identify any influence of other stakeholder downtimes on the logistic operations.

4.1.2 Important KPIs

Upon interviewing participant C, and when looking at direct warehouse KPIs in inbound, one can identify three main KPIs that are of greatest importance to inbound. These are *Dock to stock time, Labour Productivity, and Storage Accuracy;* the reason why these are the most crucial is explained by the impact they have on other KPIs and by the service level agreements (SLAs) they reflect on. For a logistics service providing firm, the ultimate goal is for the products to spend the least time possible in the system. Hence the dock to stock time is what the contractor, or online retailer in this case study, is mainly concerned with. Labour Productivity is a global KPI that applies to all warehouse activities (Staudt et al., 2015). And the ending phase of inbound, as shown earlier in Figure 4, is marked by the stocking activities. Hence the storage accuracy entails the precision of the final result of inbound operations, items being stowed on shelves and ready to be picked. Storage accuracy is measured by the percentage of items stored in their assigned shelves and racks correctly.

4.1.3 Critical Success Factors at the Case Warehouse

The CSFs of the warehouse partners are mutually agreed upon by the stakeholders. The main goal is to have smooth operations that ensure fast and accurate deliveries. Discussions in the beginning of the partnership result in SLA agreements. These mutually agreed upon SLAs are based on warehouse layout, capabilities, and other efficiencies in other locations. These values are prone to change once new data is revealed about the processes and new capabilities are found. Hence these SLAs are not they are not “set in stone”.

In this case study, these SLAs mainly concern dock to stock times that are mutually agreed upon as noted from the interviews. These are time limits defining the time allowed for a product to be in the system. The second SLA that marks a CSF for a service provider is adhering to certain efficiencies. This is measured by the time spent for each item, or seconds per item (SPI). The SPIs are defined by the contractor and each area has a different SPI, these values are determined based on the design and capabilities of the warehouse. Finally, the quality of the work done by the operators is measured by the storage accuracy in stowing. For each of these SLAs there are bare minimum values that must be delivered. Therefore, we conclude the following:

Critical Succes Factors for a service provider in the inbound part:

- Matching the allocated time limit for each product to be in the system
- Adhering to certain SPIs for the efficiency
- Achieving a predetermined quality level of storage accuracy

4.2 Modifying OEE for Warehouses

After identifying the CSFs that should be incorporated into the artefact and understanding why they are important, one must explore how these can be expressed together in a measure for OWP.

As discussed in the previous chapters, measuring OWP using the OEE metric will be explored in this research. OEE was designed for singular manufacturing equipment, hence the metric and its factors and inputs must be modified to suit the case of application in this study.

4.2.1 Warehouse Losses

As mentioned earlier, OEE is concerned with six big losses that were initially described by Nakajima. These losses linked to availability, performance, and quality are general terms for any manufacturing system. In order to formulate a modified OEE for retail warehouses, the losses of such warehouse must be identified and linked to its appropriate coefficient of OEE. Examples of these categories from the warehouse under study in this thesis will be presented below. These broad titles are aimed at serving any warehouse with similar loss definition.

The warehouse under study can be considered a combination of three systems. Firstly, the warehouse management system. A warehouse management system's primary purpose is to manage a warehouse. These systems, on the one hand, keep track of the storage capacity, or the specifications of the current storage bins (location management), and on the other hand keep track of the units that were stored (inventory management); to further optimize the storage processes, it should also include a number of control functions (ten Hompel & Schmidt, 2007). Secondly, the mechanical system that is composed of the conveyor belts, storage bins, and all mechanical equipment moving the totes throughout the warehouse. Finally, the operating partner, which in this case is CEVA. This partner runs all the operations using the WMS provided by the contractor and the mechanization partner as well.

This split of systems clarifies where downtimes can occur, there can be three downtimes that affect the availability of the overall system. These can be planned or unplanned, and both are considered when calculating the overall availability.

Availability Losses:

Downtimes concerning availability can be categorized into two major losses in this area:

- WMS downtime and mechanical downtime
- Planned downtimes and changeover between stations

Similarly, performance losses must be adapted to warehouses. These losses are more similar to the original OEE definitions than the availability losses of warehouses.

Performance Losses:

The general performance of the warehouse is reduced by two main possibilities:

- Minor stoppages in the system and idling of stations
- Slower performance or reduced speed

As for quality, since no production takes place, no defective items are made. However as mentioned earlier, storage accuracy is what measures the quality of the inbound operations.

Quality Losses:

Hence it is concluded that quality losses in the inbound part of warehouses occur from:

- Item wrongfully stored in another shelf

4.2.2 Modified OEE Framework

Figure 8 depicts the modified OEE framework synthesized in this research based on these losses. This framework is used to come up with the formulas that will constitute the modified OEE for warehouses.

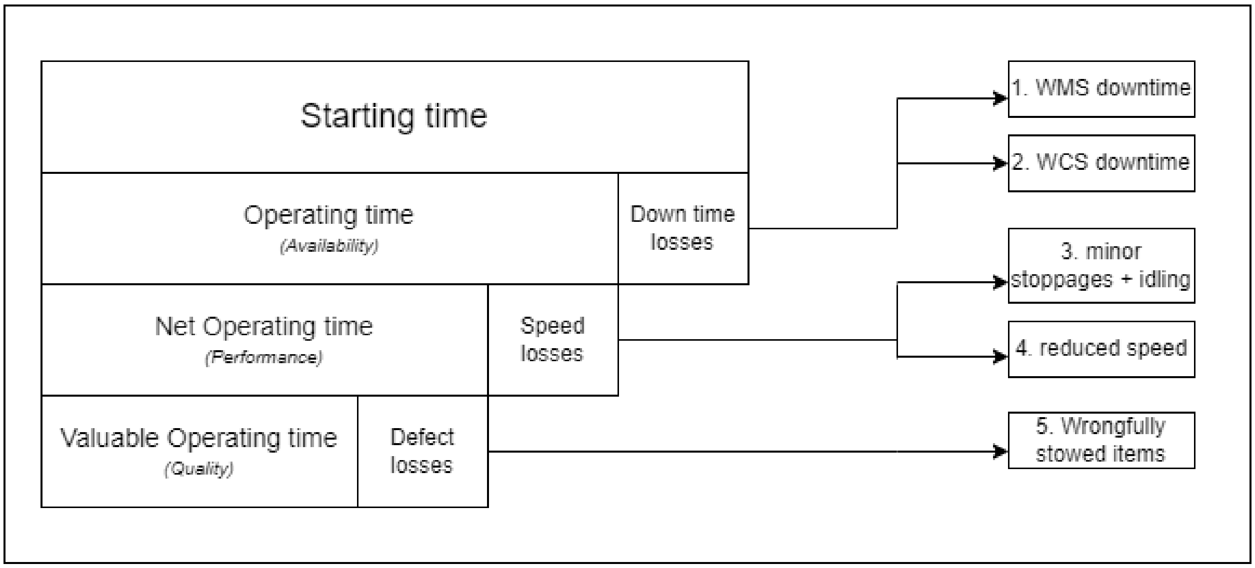

Figure 8 Modified OEE framework for warehouses (inbound)

The OEE formulation can be carried out and the coefficients can be established once the losses have been identified. The modification in this framework is primarily in the specification of these coefficients, which are what the original OEE formula uses to obtain the product of multiplying availability, performance, and quality. The beginning of this alteration was when the losses were converted to a warehouse losses. However, that alone does not achieve the goals of this study because there is still room to emphasize the CSFs even more. The losses described above showed what a warehouse perceives as compromising its CSFs, but a further change to the formula is required to also provide a ranking among these success elements.

The traditional OEE formula assumes that a change in availability has the same effect on the overall measure as that particular change in performance for example. However, in practice that may not be accurate. A five percent increase in availability does not necessarily have the same effect as a five percent increase in quality on the OWP for CEVA. This is due to the different importance of each OEE factor for each firm. For a service provider, this may be decided by the terms of payment for their services. A service provider that is paid by commission will have different priorities than that of a firm that is paid a fixed amount. Additionally, the nature of the warehouse and the items processed may make certain factors, like quality for example, more important. To take this into account, and upon the insights gathered from the interviews, incorporating weights into the calculation of the overall metric is suggested. The use of weights is further discussed in chapter 6.

To give each factor a predetermined importance, weights have been added to the three factors of OEE following one of the earliest modifications done by Raouf in 1994.

Hence the modified formulas are as follows:

$$OEE \; = \; Availability^{W_A} \text{ x } Performance^{W_P} \text{ x } Quality^{W_Q} \tag{1}$$

where,

W_A= Weight of the availability factor as determined by the management

W_P= Weight of the performance factor as determined by the management

W_Q= Weight of the quality factor as determined by the management

$$W_A + W_P + W_Q = 1 \tag{2}$$

and

$$\text{Availability} = \frac{\text{Scheduled Operating time – Unplanned downtime}}{\text{Scheduled Operating time}} \tag{3}$$

$$\text{Performance} = \frac{\text{Net Effective Operating time}}{\text{Actual Operating time}} \tag{4}$$

$$\text{Quality} = \frac{\text{Number of correctly stowed items}}{\text{Total number of stowed items}} \tag{5}$$

$$\text{Scheduled Operating time} = \text{Total available time} - \text{planned shutdown time} \tag{6}$$

$$\text{Unplanned downtime} = \text{WMS downtime} + \text{Mechanical downtime} \tag{7}$$

$$\text{Net Effective Operating time} = \text{Number of items processed x nominal SPI} \tag{8}$$

4.3 Building the Artefact

One objective of this study is to transform the modified OEE framework into a useful artefact that can be tested, possibly used in industry, or serve as a starting point for more advanced artefacts based on the knowledge discovered through this research. A spreadsheet model with the incorporated framework is built using the aforementioned formulas. This is done using Microsoft Excel for its global compatibility and ease of use. Additionally, control room personnel at the case company are familiar with the software, hence an increased research relevance is achieved. The data entered into the model are from company records. All the numbers used here are a mere representation of the excel model and how it works rather than to make implications based on the data.

4.3.1 Availability

Equations 3, 6, and 7 determine the inputs for calculating the availability coefficient in the OEE formula. Table 9 shows the input sheet that a control room specialist would look at when recording the downtimes. It starts with inserting the total available time, then from that time several losses are subtracted as shown in the modified OEE framework in Figure 8.

Availability			Time
Total Available Time			*Removed for confidentiality reasons*
Planned Shutdown (Maintenance or not in demand)			
Scheduled Operating time			
Unplanned Downtimes			
WMS downtime (min)			
Mechanical downtime (min)			
Other downtime (min)			
Total recorded downtimes (hrs)			
Actual Operating time			
Availability			

Availability (based on total time)			[redacted]

Table 9 Availability input in Excel model

It is worth noting that this table may contain more data than what the modified OEE is concerned with, this is to give flexibility in case a TEEP approach was to be taken where availability is calculated based on the total available time. In practice, the highlighted availability which is only based on the scheduled operating time is used more often. However, this may eliminate the possibility of identifying a loss concerned with too much planned shutdown. Although this difference in inclusion is not reflected in performance, a low availability based on total time can identify reduced utilization of the system. But utilizing a more specific availability, like the one provided, is preferred because one of the crucial objectives of the organization is to address unplanned downtimes.

After calculating availability, one may recalculate or reevaluate certain expected KPIs or SLAs. To put it in simpler terms, when availability is known, a more accurate prediction of number of items to be processed is achieved.

4.3.2 Performance

Equations 4 and 8 are used to calculate the performance coefficient. These are incorporated into the Excel model. However, there is more to be explained when breaking down the losses.

Performance			
# of items			[redacted]
Units per hour target			[redacted]
Nominal SPI	Cycle time in sec	29.75206612	[redacted]
Actual clocked hours			[redacted]
Net Effective Operating Time			*Removed for confidentiality reasons*
Idling Hours			[redacted]
Speed loss			[redacted]
Performance			[redacted]

Table 10 Performance input in Excel model

The nominal SPI values are mutually agreed upon by the partners, being the warehouse owner and the service provider. These rely on the warehouse's layout and capabilities as well as SPIs from other similar locations. This is changed from seconds per item to hours per item to be consistent with the time units being used. The operating time that was calculated in the availability section is not fully utilized as it is not all going into effective operating time. The net effective operating time is calculated by multiplying the nominal SPI with the total number of items processed. The reason for this reduced efficiency than the nominal performance rate is shown in Figure 8. The two main reasons being idling and minor stoppages, on top of that, reduced speed is a cause to reduce the operating time into net effective time.

4.3.2.1 Speed Loss

To calculate the speed loss, the recorded hours of idling workers and the net effective operating time are subtracted from the total operating time. The number presented as speed loss resembles the hidden impact on other KPIs. As the overall area should have processed a certain amount based on the time worked (after calculating availability) and the SPI the facility can achieve. Many operations in automated warehouses like the one being studied are not directly impacted by labour productivity when working. The degree of automation determines how much human speed has no bearing on performance levels; in a warehouse, the higher the level of automation, the more tightly controlled the performance is. Therefore, it can be concluded that the speed loss calculated can resemble the hidden impact of the downtimes as will be further discussed in chapter 6.

4.3.3 Quality and OEE Calculation

Out of all the three factors, quality is the most straightforward when it comes to inbound operations of an automated warehouse. Where it is simply the ratio of correctly stowed items in their allocated shelves to the total number of items stowed. In other words, the direct metric storage accuracy is resembled by the quality coefficient in this framework.

Quality			Units
Total # of items stowed			
# of correctly stowed items			*Removed for confidentiality reasons*
# of wrongly stowed items			
Quality			

Table 11 Quality input table in Excel

Table 11 shows its simplicity, yet it is the most crucial quality measure as not many quality KPIs apply here.

OEE Coefficients	Weight of Coefficients		Weighted Coefficients
Availability	W_A	0.5	
Performance	W_P	0.4	*Removed for confidentiality reasons*

Quality	W_Q	0.1	*Removed for confidentiality reasons*
OEE			

Table 12 OEE calculation and weights input in Excel model

After calculating all three coefficients, a display of each value and an option to input the weights for each coefficient is given. These values largely depend on the strategy of the operating firm and can be different for each service provider based on their CSFs and the type of warehouse being operated. For confidentiality reasons, the numbers suggested for the case study cannot be shared as they are derived from confidential SLAs. Table 12 shows an example how this is applied but does not suggest these particular values as weights for automated retail warehouses. Further discussion about assigning values to these weights is presented in chapter 6.

4.4 Applications of OEE in Warehousing

Several suggested uses of the model are presented here to utilize the developed artefact.

4.4.1 Tracking Performance

One of the most important functions of KPIs is to track performance over time. This enables decision makers and analysts to determine if the current strategy and way of operating is failing or not. Identifying a drop in performance over time calls for re-evaluating the current method of conducting operations.

Keeping track of OEE values and individual OEE coefficients throughout the year can help identifying trends throughout seasons. Comparing OEE values over time can show a correlation between certain events and the overall performance. Not only that but analyzing the artefact results can also pinpoint which subprocesses are performing better and when. This can be useful to help discover the reasons behind that and imitate it if possible to the lower performing subprocesses, leading to continuous improvement.

4.4.2 Continuous Improvement

A natural result of performance tracking is identifying potential areas of improvement. Analyzing times where OEE was low and identifying the most impactful coefficients follows marks the beginning of finding these areas. Studying the relationship between OEE coefficients and identifying trends may shed light into untapped potential. Furthermore, the impact of previous changes on the overall performance can be seen when looking at the records. Comparing OEE values pre-change and after the improvement can show the effectiveness of this change. Building on improvements that proved to impact the OWP will provide a more effective plan for future changes.

4.4.3 Communicating OWP Results

As a service provider, companies offer their services to potential clients. Due to the fierce competition, a competitive advantage must be met to win over these customers. Showing a good track of effective operations in current warehouses can motivate contractors to proceed with this service provider. Having a proven record strengthens the message of selling services and provides a statistical and scientific based ground to prove their effectiveness.

Additionally, as a warehouse has different stakeholders, it is important to report the performance to all business partners. The model can clarify the reasons behind falling short of forecasted results. Showing a reduced availability due to downtimes of other systems may motivate business partners to also look into solutions to increase availability.

Communicating results within company employees and logistical operators can also be useful. Showing workers how well they are performing and keeping track of OEE in real time can motivate workers further. Team leaders can also show their subordinates if their work is conforming to the quality standards or not.

4.4.4 OEE Benchmark and Thresholds

According to Nakajima, under ideal conditions, organizations should have greater than 90 percent availability, greater than 95 percent performance, and, surprisingly, greater than 99 percent quality. These figures translate into an OEE of 85 percent for world-class firms, which Nakajima considers to be a good benchmark for typical manufacturing capability.

However, these values were derived from a manufacturing context. Nevertheless, they can set a general guideline for this model as no data on OEE in warehousing is available in order to set benchmarks. In practice, bad days exist where operations have a major downfall. These thresholds can indicate the limits where the performance of the service provider is acceptable.

Daily meetings occur at the case company where the performance of the previous day is discussed. Many KPIs are involved in this discussion and often a scope of investigation is not defined. Discussing OEE coefficients and finding reasons for their fall beneath acceptable thresholds can take place with the help of this artefact.

Communicating extremely high downtimes to responsible stakeholders should be done when availability falls short of the threshold. Availability is the least controllable coefficient by the service provider therefore effective and accurate communication must take place with the help of recorded OEE data.

Performance falling behind thresholds is of great importance for service providers. It resembles how well the job is being done. Once a drop is noticed, immediate attention to team leaders must be given to identify any reasons for reduced performance. Idling workers may be noticed, and slow operations can be changed in time to prevent the low performance rate lasting for long.

The highest and most sensitive threshold is quality. Particularly for service providers, extremely high thresholds are made by contractors in SLAs. If quality of stowing is low, it affects picking which is the most important and costly process in a warehouse. Therefore, possible actions to be taken when quality is low include increased training of workers. Additionally, awareness of the impact of low-quality levels can be raised to further reduce errors.

The utility of the artefact is evaluated by experts as shown in the next chapter. Practical implications and further discussion of the applications of OEE takes place in the discussion chapter.

Chapter 5: Evaluation

As mentioned earlier, an analytical and descriptive evaluation is carried out for this study. The artefact was presented to experts in the field from the case study, to evaluate its usefulness and utility amongst other criteria. Semi-structured interviews were used to present relevant results and give experts more room to evaluate based on their grounds and needs that align with the DSR approach, to achieve the goal of solving a real-life problem. Furthermore, the artefact was evaluated using these interviews to obtain a more detailed evaluation and to allow for recommendations and future research suggestions. For a similar study in a new field, it is argued that a qualitative evaluation of the artefact based on expert opinions should come before quantitative evaluation. This is due to the exploratory nature of the study, and that there has been little research on applying OEE to warehouses. Appendix C contains the template used to conduct these interviews where three employees from different organizational divisions were interviewed.

The artefact was evaluated for its complexity, applicability in practice, relevance to industry, and its utility. Interviewees were given room for suggestions to improve the artefact and concerns about the model. In the following subsections, these headlines are outlined and elaborated.

5.1 Complexity of the Artefact

One of the major criteria for evaluating the artefact is its complexity. The artefact developed was made as easy as possible. Too much complexity reduces its usability to begin with and hence the artefact renders useless. On the other hand, if the model is oversimplified then the utility and development of this artefact is questioned. The first question asked to the interviewees after gathering basic information is their opinion on the complexity of the artefact. Throughout the interview more detailed questions regarding the difficulty of use, familiarity of terms and clarity of fields are asked. Table 13 shows the summary of the evaluation for complexity from the interviews.

5.1.1 Initial Impressions

The overall initial impressions were good. The split between the three coefficients and the three areas of inbound made it logical. Color coding was complimented and described to make certain numbers “pop out”. Participant B was thrown away by the amount of numbers at the first glance but after viewing the model for a minute he understood the model more and things were clear for him. One important matter that had to be verified is its complexity for the people that will input the data into the model particularly. It was expressed by the control room manager that the model is readable for someone in the control room after a brief introduction to some operational terms. Participant C described the artefact to be well structured and can be easily navigated through.

5.1.2 Difficulty of Use and Familiarity

The difficulty of use was evaluated to make sure that if implemented, not many issues would arise due to perceived complexity by the user. The evaluations of the participants indicated that it is important to identify descriptions of unfamiliar terms as was done in the artefact. Therefore, it is important to note that if the terms used in an artefact are different to those that the user is familiar with, descriptions should follow these unfamiliar terminologies to make it easier to use. Participant C was the most familiar with the model as data and business analytics is her field. She expressed familiarity with the terms used and that it is not too complex.

5.1.3 Possible Confusions and Final Impressions

Structuring the artefact in a logical manner was helpful to clarify the numbers being calculated. Initially it was a little bit confusing for participant A due to the split between the availability table and what came after it, the performance and quality tables joined together. Most of the confusions that were clarified were due to the visual presentation of the artefact. It is vital to design the artefact in a way that is user friendly and self-explanatory. The sequence of the fields helped to give a clearer understanding of the numbers and what the calculations mean. Participant C raised the issue of a confusion that may arise to the users inputting the data, she mentioned that some fields may be prone to subjectivity.

The overall final impressions were positive and showed a better understanding of the model presented. It seemed logical and straight-forward for all the participants after shortly navigating through the artefact.

Question Area	Participant A	Participant B	Participant C
Initial impressions	Logical and clear split of inbound activities	Initially overwhelmed by the numbers	Well structured
Difficulty of use	Simple yet just enough complexity	Readable for control room specialists	Not difficult
Clarity of fields	Mostly clear	Clear	Clear
Familiarity of terminologies	Noted one different terminology from practice	Familiar	Familiar
Possible confusions	Setup of tables; quality of unloading and receive	Visuals can stress certain numbers	Possible subjectivity of users filling in model
Overall impression	Works well	Reads through easily; needs more appealing visuals	Structure looks fine

Table 13 Complexity of artefact evaluation

5.2 Applicability and Relevance of the Artefact

Another important criterion for evaluating the artefact is its practicality. As previously stated, the primary reason a DSR approach was chosen was to provide a solution to a real-world problem. As a result, it is critical that the artefact can be used in practice to solve or contribute to the solution of that problem. Additionally, relevance of the artefact must be achieved to ensure its applicability. Table 14 shows the overall evaluation for these criteria.

5.2.1 Inputs of the Model

For the artefact to be applicable, the data to be input into it must be present. Interviewees were asked if this data is measured in practice. The responses were positive, the inputs are in fact available, and the service provider does measure it. However, it was noted by participant B that not all measurements are made by the service provider itself, WMS downtimes are measured using the ticket system provided by the contractor i.e. the online retailer owning this warehouse. Hence one should ensure that the model being developed can be used in practice and that the data needed for the results are not missing. Participant C particularly criticized the lack of objectivity in obtaining the inputs of the model. Having reliable data input to the model is important and it is crucial to make it least subjective as possible.

5.2.2 OEE Components

Specific questions about OEE related components were asked to evaluate the relevance of the modified OEE to the warehouse industry.

5.2.2.1 Warehouse Losses

The modified OEE can only be used as a measure of OWP if the three components of modified OEE are relevant to the case study. The reactions towards defining the losses of warehouses as presented in the artefact were positive. Interviewees had a good understanding of availability, performance, and quality and after a brief navigation throughout the artefact, their components were clearly obvious and confirmation that it was relevant was made.

5.2.2.2 OEE Calculation and the Use of Weights

Questions about the proposed method of calculating OEE using weights were raised to the different interviewees. The responses were not all the same. Participant A entertained the idea of giving different factors different weights. He added that some areas of the warehouse may have different weights than others as he stated that quality in the outbound part of the warehouse is more important than other parts of the warehouse. Participant B, the control room manager, made it clear that he believes all coefficients should be treated equally. He elaborated on the fact that they are all important and that one coefficient cannot be overlooked. The argument of setting weights based on the payment of the contractor for each factor was made to the participant, yet he stood by his initial opinion that even if the markup is different, they should all be given the same importance. Participant C, however, argued that using weights should be based upon the mark up or payment structure by the contractor. It is a common practice that service providers are paid more money for some SLAs than others. She stressed the importance of maintaining quality and mentioned that the minimum threshold set is really high.

5.2.2.3 OEE as an OWP

Finally, participants were asked if they understood what OEE measures. While all of them understood the concept and were familiarized with the formulas, participant A mentioned that reporting each factor separately is needed more than the overall measure. The rest, however, agreed on the importance of the overall measure more and expressed some utilities where they think it will be useful. Overall, the concept of using the modified OEE as measure of OWP and their company's performance was accepted. Participant C strongly believes that using OEE gives a total representation of your site's performance and mentioned that it gives an indication if their performance is growing or not.

5.2.3 Matching Expectations

The participants had different views based on their organizational function. This was evident when they were asked if the artefact met their initial expectations of an overall measure of their performance. Participant A clearly identified that this is what he was expecting as he had a good background about the OEE concept. On the other hand, participant B had a different view. Due to the lack of knowledge of OEE, it took him a while to totally grasp the idea and mention that it is suitable to be an overall measure. Participant C mentioned that her expectations of a model to measure overall performance was different. The expectation was to use OEE, however she argued there needed to be more elaboration for the detailed components of the coefficients. The basis of the artefact however was agreed upon by this participant.

Question Area	Participant A	Participant B	Participant C
Inputs of the model	Existing and measured	Existing and measured	Existing and measured
Warehouse losses	Relevant	Relevant	Relevant
Use of weights	Agreed	Disagreed	Agreed
OEE as an OWP measure	Agreed	Agreed	Agreed
Matching Expectations of an OWP	Yes, more details needed	Yes	Good basis to build upon
Overall impression	Applicable	Applicable	Applicable after further development

Table 14 Applicability and relevance of artefact evaluation

5.3 Applications of OEE in Warehousing

When asked about the utility of the presented artefact, the interviewees mentioned a few uses where they think the developed artefact can provide some utility. One use that was common between all of the interviewees was using the model to report to customers (warehouse owners) all the necessary data.

Table 15 summarizes the interview results in evaluating the utility of the presented artefact.

5.3.1 Reporting to Stakeholders

The first utility mentioned by all the participants regarded reporting the service provider's performance to the customer. As mentioned earlier, there are three main system components, the WMS, mechanical system, and the operating service provider. All participants agreed that it is crucial to justify any lack of performance or shortcomings from forecasted results. Participant A believes it can be used in daily communications with customers to "explain what's going on our floors". Participant B stated that "I think it would be very useful not only for ourselves, but also reporting it to different stakeholders". He mentioned that team coordinators can use it to show the operators how the quality is within each different process. Participant C also identified the importance of reporting such data to partners and that simple overall numbers are useful for that.

5.3.2 Tracking Performance

Another suggested use was to track the performance of the company over time. As they agree that these coefficients are relevant to the overall company performance, it was suggested to keep track of all coefficients as well as overall OEE. This is useful for showing the difference between the performance of a warehouse in ramp-up phase and after it has been live for a while. Not only that, but it can be used to measure the impact of a change in equipment or training for example on the overall performance over time. Participants B and C pointed out this utility clearly.

5.3.3 Continuous Improvement

Participant C stressed the importance of using the components to identify areas of improvement. She argues that there is an effect for each coefficient not only on OEE, but there is a hidden impact of downtimes on performance. Identifying this impact is an area of continuous improvement that the company must invest into. Also identifying if an improvement was reflected upon in the overall performance as mentioned earlier was said to be a useful way of knowing if further improvements in that direction are "worth it".

5.3.4 Comparing Performance to Other Warehouses

All participants mentioned that this indicator can be used to draw comparisons between warehouses operated by their company. However, when discussing this further, participant A mentioned that it will be interesting to compare different warehouses using this artefact. However, he raised the issue that they may operate differently or have different specific definitions. Not only that, but the design of the warehouses is not the same, hence it is interesting yet tricky. Participant C had the same thoughts and mentioned that other warehouses may measure things differently, especially downtimes. Therefore, she also thinks that it will be challenging to make it a universal artefact without further development.

Utility	Participant A	Participant B	Participant C
Reporting to Stakeholders	Mentioned	Mentioned	Mentioned
Tracking Performance	Mentioned	Mentioned	Mentioned
Continuous Improvement	Not mentioned	Not Mentioned	Mentioned
Comparing to other warehouses	Tricky	Mentioned	Tricky

Table 15 Artefact utility evaluation

5.4 Suggestions and Concerns

At the end of each interview, participants were asked to suggest improvements to be made and mentioned any undiscussed concerns. There were two main concerns that were expressed. These regarded the user interface and visual aspect of the artefact; and making the artefact more in line with the current system by improving the subjectivity of measurements and interactivity of the model.

5.4.1 Improving User Interface

Although participant A liked the color coding and the overall structure of the artefact, he suggested making some improvements by simplifying some of the tables. He stated that for some users who may input information in the model, there may be a separate simplified page with less fields and data in it. In other words, splitting the model into a simpler one and a more complex one where the unexperienced users input data into the simple format only. Participant B stressed the importance of improving the visual appeal to the interface. Suggestions were made to make separate tabs for overall results to be presented in a more visually appealing manner. Having a split between a more complex page and a simpler one was also suggested by participant B.

5.4.2 Connecting to System Interactively

When asked about improvements to be made about the artefact, participant C instantly stated an interactive and less subjective model. Designing the artefact in a way that measures the actual downtime and translates it into impact on performance. She stated that it is desired to eliminate subjectivity of the users filling in the model as much as possible. An emphasis was made on having an interactive model that takes in the downtimes and outputs the impacts of these downtimes. It was understood that this is one of the most important utilities of the expected model that was pictured by this participant.

5.4.3 Input Data

Participant A suggested using different weights for each coefficient for each activity rather than using the same weights for all subprocesses. He expressed that in the outbound area quality is given a higher importance as the product is approaching the customer thus more attention should be given to quality. Another concern that was raised about the input data to the model was that different stakeholders may have different definitions of downtimes as was expressed by participant B. He mentioned the importance of aligning the downtimes together from all the stakeholders' perspectives. Differentiating between what the service provider counts as a downtime and what the mechanical partner considers a downtime is crucial especially when reporting to business partners like the contractor. Participant C highlighted the fact that downtimes should not be counted twice if they overlap. She said that this would give inaccurate conclusions based on the results of the model. A suggestion for improvement was made where it is automatically configured that if both downtimes occur simultaneously that the earliest downtime start time is what should be considered and the last one to be resolved is to be considered the ending point of this downtime.

5.5 Quantitative Analysis

As mentioned earlier, it is argued that qualitative evaluation must precede any quantitative evaluation for a new model like the developed artefact. As the artefact is new, its relevance and applicability must be tested with experts in the industry their input is to be considered before proceeding to the next step of validating the model with numbers.

Two quantitative evaluations took place for the model, firstly a sensitivity analysis is performed using Excel to identify how OEE values change with incremental changes of availability and performance. Secondly, a test for correlation between OEE and dock to stock times is performed using JASP software.

5.5.1 Sensitivity Analysis

The first step of evaluating the model quantitatively is to test how sensitive the output is to changes in the input of the model. This sensitivity analysis can reveal if a significant change in an input has a significant effect on the output and by how much. The inputs of the OEE formula are the availability, performance, and quality coefficients. However, as mentioned earlier, the quality thresholds are above 99% meaning that the change in OEE of fractions of 1% may be negligible. Therefore, in this specific case study, a sensitivity analysis of performance and availability against OEE values takes place.

The calculations of the model assumed weights values for performance and availability coefficients. With the availability weight at 0.5 (WA=0.5) and a performance weight at 0.4 (WP=0.4). This leaves the quality coefficient weight at 0.1 (WQ=0.1). Table 16 shows the results of the sensitivity analysis. It can be noted that a 15% increase in availability at 75% performance yields in an 11.94% increase in OEE. Another example would be a 15% increase in performance from 75% to 100% at a 90% availability rate, this yields in an increase of 10.32% in OEE. This shows that OEE does in fact significantly change with an increment relatively close to the input change.

		Performance										
	OEE	75.00%	77.50%	80.00%	82.50%	85.00%	87.50%	90.00%	92.50%	95.00%	97.50%	100.00%
Availability	75.00%	77.19%	78.21%	79.20%	80.19%	81.15%	82.10%	83.03%	83.94%	84.84%	85.73%	86.60%
	77.50%	78.46%	79.50%	80.51%	81.51%	82.49%	83.45%	84.40%	85.33%	86.24%	87.14%	88.03%
	80.00%	79.72%	80.77%	81.80%	82.82%	83.81%	84.79%	85.75%	86.69%	87.62%	88.54%	89.44%
	82.50%	80.95%	82.02%	83.07%	84.10%	85.11%	86.10%	87.08%	88.04%	88.98%	89.91%	90.83%
	85.00%	82.17%	83.26%	84.32%	85.36%	86.39%	87.40%	88.39%	89.36%	90.32%	91.26%	92.19%
	87.50%	83.37%	84.47%	85.55%	86.61%	87.65%	88.67%	89.68%	90.67%	91.64%	92.60%	93.54%
	90.00%	84.55%	85.67%	86.76%	87.84%	88.89%	89.93%	90.95%	91.95%	92.94%	93.91%	94.87%
	92.50%	85.72%	86.85%	87.96%	89.05%	90.12%	91.17%	92.20%	93.22%	94.22%	95.20%	96.17%
	95.00%	86.87%	88.02%	89.14%	90.25%	91.33%	92.40%	93.44%	94.47%	95.49%	96.48%	97.46%
	97.50%	88.01%	89.17%	90.31%	91.43%	92.52%	93.60%	94.66%	95.71%	96.73%	97.74%	98.74%
	100.00%	89.13%	90.30%	91.46%	92.59%	93.70%	94.80%	95.87%	96.93%	97.97%	98.99%	100.00%

Table 16 Sensitivity analysis of Excel model

5.5.2 Correlation Test

To evaluate the performance of OEE as an OWP measure, a test for correlation between OEE values and dock-to-stock scores took place. The results are discussed in chapter 6, however it is worth noting that the data collected is limited and to arrive at more relevant results more data points across months of operations must take place to increase confidence levels of the conclusion of this correlation test.

A Bayesian correlation test using JASP was done and the results are presented. Firstly, the OEE values and the corresponding dock-to-stock scores represent the data input for the JASP analysis. Dock-to-stock times cannot be shared publicly and are represented by aggregated scores for publication purposes. Appendix D contains all the recorded data and the filled Excel models upon data collection from the warehouse. Table 17 shows the final results and present the OEE values for the data collected and their corresponding aggregated dock-to-stock scores. This was used as the input data for the JASP software to test for Bayesian correlation. This method is chosen due to prior belief that the two values are correlated and hence they hypothesis is tested based on that. This leaves a null hypothesis (H_0) that OEE and dock-to-stock are not correlated, and an alternative hypothesis (H_1) that there is in fact a correlation.

Day	OEE	Score
Monday	105.45%	0.021739
Tuesday	105.67%	0.041667
Wednesday	100.84%	0.027778
Thursday	115.52%	0.019231
Friday	102.16%	0.02381

Table 17 OEE vs Dock-to-stock scores

Table 18 results summarize the JASP output.

Bayesian Pearson Correlations

Variable		OEE	Score
1. OEE	n	—	
	Pearson's r	—	
	BF_{10}	—	
2. Score	n	5.000	—
	Pearson's r	-0.332	—
	BF_{10}	0.609	—

* $BF_{10} > 10$, ** $BF_{10} > 30$, *** $BF_{10} > 100$

Table 18 Bayesian correlation test results

To interpret this JASP output, two values must be evaluated. These are the Pearson's r and Bayesian Factor. The Bayesian Factor indicates how strong the evidence is of an existing correlation. Additionally, the Pearson's r indicates the strength of the correlation and its direction, Table 19 shows the interpretation table of the Bayesian Factor values and Table 20 gives insight on how to interpret Pearson's r values when testing for correlation.

BF_{10} Evidence for H_1		BF_{10} Evidence for H_0
1	No evidence	1
1-3	Anecdotal evidence	1-1/3
3-10	Moderate evidence	1/3-1/10
10-30	Strong evidence	1/10-1/30
30-10	Very strong evidence	1/30-1/100
>100	Extreme evidence	< 1/100

Table 19 Bayesian Factor values interpretation, Source: (Tona et al., 2019) *page 8*

Correlation Coefficient Value (r)	Direction and Strength of Correlation
-1	Perfectly negative
-0.8	Strongly negative
-0.5	Moderately negative
-0.2	Weakly negative
0	No association
0.2	Weakly positive
0.5	Moderately positive
0.8	Strongly positive
1	Perfectly positive

Table 20 Pearson's r interpretation, Source: (Ratnasari et al., 2016) *page 2*

Moreover, a scatter plot presented in Figure 9 shows a visualization of the correlation test results.

Bayesian Correlation Pairwise Plots:

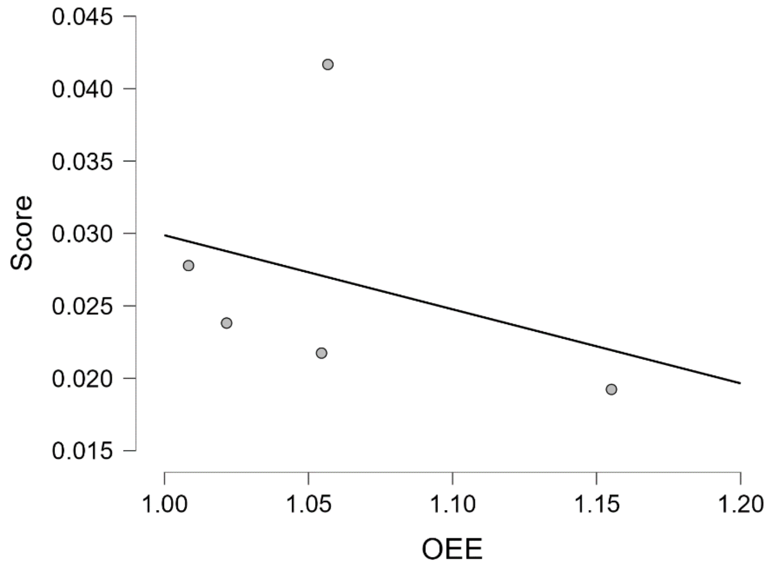

Figure 9 OEE and dock-to-stock scatter plot

Based on the results, and due to the low value of the Bayesian Factor in Table 18 the alternative hypothesis cannot be verified and there is reason to accept the null hypothesis. These results suggest a negative correlation based on the Pearson's r which is opposed to the expected positive correlation before testing the results. Further discussion on these results and reasons for the differences is presented in the next chapter.

Chapter 6: Discussion

In this chapter, further discussion about the artefact and its utility is presented. Key findings of the research will be identified and elaboration on certain important points is made. The evaluations are interpreted and discussed, and limitations of this research are also mentioned. Recommendations and reflections upon this research are discussed to provide more insight of the direction of this research and how it can be followed upon in future studies or applications.

6.1 Evaluation Results

As mentioned earlier, several criteria were discussed in the interviews to evaluate the artefact. In this subsection, a broader discussion takes place on the results of these interviews.

6.1.1 Complexity of the Model

When evaluating the model, it is important to evaluate it based on the perceived objectives. An overall measure should not be too complex. The primary goal of having an overall measure is to eliminate the confusions and complications that may happen when assessing numerous KPIs. The evaluations from the expert interviews showed that the presented model was logical and well structured.

Another important aspect of complexity that needs to be considered whenever designing such model is the difficulty of use. It is crucial that the users of any model are easily familiarized with it to avoid any mistakes happening due to misconceptions. The results were positive in this aspect which indicates that the simplicity of a model is needed in this area. It is argued that the usefulness of an artefact may be largely affected by its complexity. A very complex model, even if it provides relevant results may render useless if it cannot be used or if it is used wrongly.

Using familiar terminology and providing descriptions for cells that might be confusing helped achieve this simplicity and ease of use. Matching the metrics used in the field with a model should be considered when designing an artefact. This helps bridge the gap between scientific research and practice, helping workers from the industry to use models developed in the literature.

6.1.2 Applicability and Relevance of the Model

Another vital aspect of any new work is its applicability and relevance to the field. Evaluation in this direction was made to ensure that further research in this direction is useful. Defining the losses of warehouses was the initial step of modifying OEE to the warehousing industry. The relevance of this aspect of the model is crucially important for this research to be generalizable. Expert interviews identified that the losses used to calculate OEE are relevant to the field and are inclusive of important KPIs and the CSFs of the firm.

However, it must be clarified that the model presented was developed for the inbound part of the warehouse. This is worth noting because the defining KPIs of the quality coefficient are different for each warehouse activity. As was pointed out by the interviews, the levels of quality of different activities vary in importance and in measurement methods. Therefore, when expanding this model further to other warehouse areas, redefining quality losses is required.

Building on that idea, one should be aware that the nature of the warehouse and the items processed in it can also redefine quality losses. The items processed in this case study were that of an online retailer in the clothing industry. But if the items are of fragile nature, other quality related considerations need to be made. Another thing to be considered is the stage of the warehouse operations. The warehouse under study was in a ramp-up phase, when extending this model to another case for example, the fluctuations in OEE values may be different. This affects the ability of OEE to be considered as a benchmarking tool to compare warehouses. Nevertheless, the losses mentioned in this study can be generalized to all automated warehouses but comparisons of OEE values may be challenging across different warehouses in different stages.

6.1.3 Utility of OEE

In the conducted interviews, participants expressed multiple uses for the artefact. Firstly, the use that all interviewees mentioned as an initial response was using this number in reporting to other stakeholders. While each participant focused on different stakeholders, it was clear that having an overall number representing the service provider's effectiveness is desirable when communicating advancements in business. Other uses such as tracking OEE over time and continuous improvement were mentioned by several participants. However, it is worth noting that the participants were not keen for all the utilities in the same way. Not only that, but each interviewee expressed how they would use this number to satisfy their needs in their department. To clarify, the stakeholders mentioned by the manager of business analytics were different to that of the control room manager. The former is more in contact with the contractor whereas the latter is concerned about interactions with other business partners more.

Further discussion on the use of OEE in practice and certain implications that need to be addressed will be presented in the next subsection.

6.1.4 Quantitative Analysis

The sensitivity analysis shown in Table 16 suggests that the model presented has reasonable sensitivity and that OEE values change relatively closely to the input increment changes. It wis worth noting that one of the reasons for this analysis is to verify the suggested use of weights and their values. At weights of 0.5 and 0.4 for availability and performance coefficients respectively, the model seems to provide relevant results. Hence a suggestion to the industry for using these weight values in practice for similar case studies is made.

Regarding the OEE calculations using real data from the warehouse, it is noted that performance coefficients often were more than 100%, rendering OEE values of greater than 100%. The reason behind this is an overestimation of the nominal SPI that is possible for operation. Leading to a lower target units per hour than is feasible in practice. This is evident when the number of items processed is significantly greater than forecasted amounts. Suggestions to reevaluate nominal SPIs is made to showcase the true potential of the warehouse operations.

Moreover, the performance coefficients were the least realistic showing that it is of most importance to arrive at more accurate SPIs to get a realistic assessment of the warehouse performance. It must be noted that this applied to unloading and receive activities, whereas stowing activities had realistic performance values.

To validate the use of OEE as an OWP measure, a correlation test was done to test the correlation between OEE values and dock-to-stock scores. The results showed that there is no evidence for a correlation between the two values, hence the null hypothesis has not been rejected. Reasons for this result that is opposed to the expected result of a positive correlation include the inaccuracy of nominal SPI values. The overestimated SPIs result in significantly higher performance coefficients, which lead to OEE values above 100% which should not be the case. Moreover, the data collection happened at the very last phases of the project and due to time limitation not enough days' data was collected and analyzed, further research into comparing values from this model with dock-to-stock scores must take place. Upon validating the performance of OEE as an OWP measure, further analysis into its performance as an OWP as opposed to DEA models can take place.

6.2 Practical Implications

Naturally, a performance measure is used to evaluate the firm's activities. However, when addressing the uses of the measure provided using OEE, some issues might be questioned, and some areas may seem vague. In the following subsections, the practical implications of using a modified OEE model in warehousing are raised.

6.2.1 Addressing Areas of Improvement in the Warehouse

There are many ways to assess reasons for failure and look for solutions. One of the uses that were mentioned for using the model is to find where the problem occurs and how impactful it is on the overall warehouse performance. While the model presented intends to calculate OEE, it can be used to identify which of the coefficients was the most impactful on the overall performance. Furthermore, the components of this coefficient can be looked at to report where the biggest loss is coming from. For example, if the calculated OEE had a five percent reduction from the previous day, one can identify if there is only one coefficient affecting it and why was that value less than usual.

Although the built artefact is built from a service provider's perspective, it is more inclusive of the problems that arise in a warehouse. If presented to contractors or other stakeholders, the model can show the contribution of each stakeholder to the losses. To clarify, splitting downtimes into a WMS downtime and a mechanical downtime, sheds more light into which area is causing more reduced availability.

6.2.2 Performance and Quality Tradeoff

Another issue of concern is increasing one coefficient at the cost of the other. Usually, this happens between performance and quality. Increasing performance speed blindly may result in quality losses. The traditional OEE formula negates this relationship by multiplying the scores together. If performance is increased by ten percent at the cost of reducing quality by the same amount then then the overall product is not affected. Moreover, reduced quality can have an impact on the proceeding outbound operations. A reduced performance in picking activities will be caused if the items are not stowed correctly, leading to a domino effect of a reduced inbound quality hence impacting the OWP of the entire warehouse.

However, in the proposed model of using weights for the coefficients, this negating effect decreases. Hence in theory, one can increase performance at the cost of reducing quality significantly without a significant impact on the overall measure. This may be misleading and be inaccurate results may be manipulated. To prevent this, a significantly high bare minimum threshold of quality is made that the service provider must adhere to in order to get the markup on the quality factor.

Additionally, the three coefficients are often compared separately when tracking performance over time for example. This provides a more objective comparison and promotes improving all aspects of OEE rather than focusing on one.

6.2.3 Comparing Warehouse Performance

The model enables firms to compare their performance over time withing the warehouse. However, as noted by participant C, comparisons amongst other CEVA managed warehouses is possible. One factor to consider, however, is to what extent the cases being compared are similar. Product types and automation levels are important factors to consider. Quality levels at manual warehouses with less automation will be significantly lower. The current case study deploys a state-of-the-art WMS that assigns items to precise shelf locations. However, another warehouse may not have the same IT facilities which results in a lower storage accuracy level. Moreover, the stage of warehouse operation must be considered. Comparing similar warehouses may be possible if the records from the ramp-up phase from another warehouse are compared to the real time performance of the current warehouse in this case study.

It was expressed by case employees that the model generated is a generalizable model that can be deployed in other partner warehouses. While this may be true to an extent, attention to similarities of the compared warehouses must precede any comparisons.

6.3 Scientific Implications

The following subsections discuss some findings that have a theoretical implication regarding the application of OEE in warehousing.

6.3.1 Interdependency of OEE Coefficients

It was mentioned earlier that one of the reasons OEE was chosen as an OWP measure is that three factors can be regarded as stand-alone factors independent from each other. While that is true by the definitions of formulas, there may be some correlation between OEE factors.

6.3.1.1 Downtimes Impacting Performance

While by definition the coefficients of OEE are independent of each other, there may be a hidden impact of downtimes on performance. The performance efficiency should be independent from the availability rate in OEE. This issue was raised by multiple employees throughout the case study. When an error is fixed, operations do not always go back to normal immediately after the downtime is over and the error is fixed.

Performance is defined by the ratio of the actual number of items processed, to the items that could have been processed at a nominal SPI. This is based on the amount of time available after all downtimes have been deducted. Therefore, by definition, a downtime should not affect the performance coefficient of OEE. Figure 10 shows a visualization of this hidden impact. The curve shows that performance gradually increases back to normal rather than a sudden jump to nominal speeds.

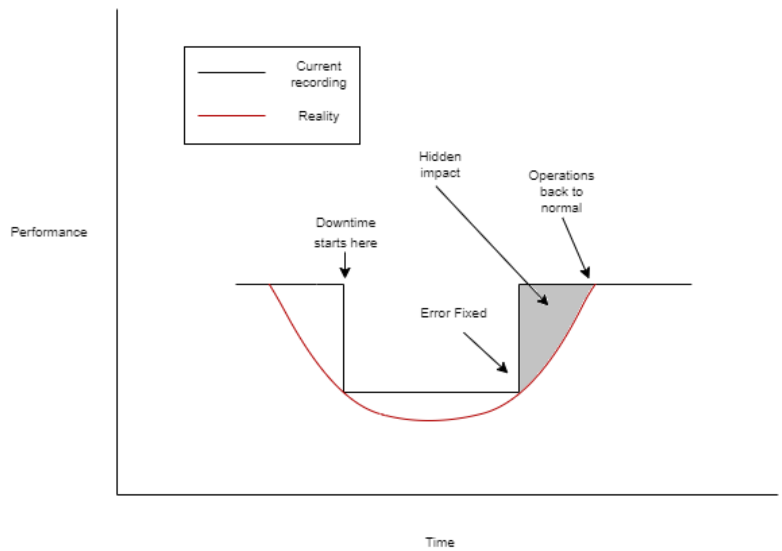

Figure 10 Hidden impact of downtimes

As mentioned earlier in the report, the speed loss that the model calculates when calculating performance can be a resemblance of this hidden impact. This is more valid for automated operations as the variability in performance speeds is less due to automation.

Another reason why there may be a hidden impact of downtimes regards moving the workers from a process that is unavailable due to an error in WMS for example. While this error is being fixed, operators may be moved to another warehouse area. If employees from the receive subprocess of inbound are sent to work in stowing, a slower performance will arise due to the lack of training for these operators on stowing activities. Not only that, but this will also result in more errors by these untrained workers.

Additionally, when an error is fixed, operations do not always start at the same time. Following upon the previous example, if the error at the receive subprocess area is fixed, the workers still need to be informed and travel from the stowing area to the receive facility. The travel time between warehouse facilities after an error is fixed is also regarded as a hidden impact resembles part of the speed loss value determined by the model.

6.3.2 Using OEE as a Benchmark

While most of the uses that were mentioned are validated in the literature, other uses that were suggested are of concern to some researchers.

There has been much discussion over the years about the definition and application of OEE. Many practitioners have discovered that it has multiple uses and definitions, causing significant confusion when comparing machines or companies. Unfortunately, OEE was not intended to be used to compare machines, plants, or businesses, but it has evolved to these common levels of misuse. Although OEE is not a statistically valid metric, it has long been used as one (Williamson, 2006). Using OEE as a benchmark to compare performances with others has been criticized by some and approved by others.

Dal et al. (2000) mentioned that OEE can be used with an internally focused benchmark where an organization sets improvement objectives, or it can be used with an external target set by the father of TPM (Nakajima, 1988) as an OEE figure of 85%. In a different field, other researchers used OEE to benchmark efficiency of port terminals was done by such as Pinto et al. (2017).

Identifying optimal OEE figures and comparing OEE between firms or shops is difficult due to different definitions of OEE and other differences between companies. Some authors, however, have attempted to do so. According to Nakajima, under ideal conditions, organizations should have greater than 90 percent availability, greater than 95 percent performance, and, surprisingly, greater than 99 percent quality. These figures would result in an 85 percent OEE for world-class firms, and Nakajima considers this figure to be a good benchmark for a typical manufacturing capability. This figure corresponds to Ericsson’s (1997) summary of various OEE measurements, in which OEE ranges from 0.30 to 0.80. These disparities highlight the difficulties in comparing OEE across processes, factories, and manufacturing sectors. (Bamber et al., 2003)

These concerns align with the possible difficulties that the interviewees mentioned when trying to compare different sites using this model. This is due to the different types of items processed in other warehouses and different operation stages as well. For example, a warehouse in a ramp-up phase has different CSFs than a warehouse that has been fully running for a while. Furthermore, warehouses operating with more fragile items may have different quality defining losses for OEE. For these reasons and other differences in warehouse types, using OEE as a benchmark to compare all warehouses to each other is not yet applicable.

Further research into this field to compare the usefulness of OEE as a benchmarking value should be done as the literature found in this area is scarce.

6.4 Recommendations and Reflections

The model presented in this research cannot be considered as a final solution to the use of OEE as an OWP. A modified OEE model was developed, and a suggested artefact was built using the developed warehouse OEE framework. Upon evaluation and during the study, thoughts and reflections on future directions were noted. The following subsections discuss recommendations based on this research and possible future research building on what this thesis has presented.

6.4.1 Recommendations for the Industry

A model was built for measuring a modified OEE for the inbound parts of the warehouse. Implementing this in practice and verifying its utility marks the initial recommendation for utilizing this research. As the evaluations suggest, the artefact can be tested and re-evaluated to make further developed iterations based on the presented artefact and OEE model. After verifying the utility of it using quantitative analysis, further adaptations can be made to explore other warehouse areas.

Following a similar methodology is advised for its practicality and ability to maintain relevance to the industry.

6.4.1.1 Using OEE as a measure of Improvement

Currently, the case company tracks KPIs over time for numerous KPIs. Tracking one OWP measure over time while maintaining more detailed inputs to provide further insight is advised. While OEE is useful as an effectiveness monitor, its true potential is revealed when used as a measure of improvement. If this metric is used to manage improvement rather than just monitor it, it will provide a useful guide to aspects of the manufacturing process where inefficiencies can be targeted (Dal et al., 2000a).

6.4.1.2 Improving Data Accuracy

The inputs to the model are essentially what decide the value of the final overall measure of performance. Ensuring high accuracy of downtime recordings, working hours, labour idling hours, and nominal SPIs is crucial for the effectiveness of this measure. A more concise and objective manner that is agreed upon by the stakeholders to define downtimes should be aimed for. Justifying this with the impact of downtimes on performance and the OEE is provided by the model and can be a good foundation for arriving at such conclusion.

The research showed that inaccurate nominal SPI values can result in unrealistic performance measures. Reevaluating target units per hour rates for unloading and receive activities is advised for the case company to arrive at a more realistic performance assessment.

6.4.2 Limitations and Further Research Directions

Since OEE has not been explored in warehousing yet, this research's main challenge was coming up with a newly modified OEE framework. Due to the nature of the study and the time limit, limited quantitative analysis of the model's performance was made. DSR allows for further iterations until a fully satisfying artefact is achieved or the allocated time and budget is used. Additionally, a single case study research may limit the external validity of the results.

Further research into validating this model with other case studies increases the generalizability of this model. Moreover, more data points to test for correlation between OEE and dock-to-stock scores is advised for further quantitative analysis. In this thesis, a single case study with an industry leader was done and it was assumed that external validity can be achieved as they set the standards for the rest. However, upon reflecting on the case study, there are many more types of items and warehouses that may require different definitions of warehouse losses and OEE coefficients.

The warehouse under study did not have the issue of damaging items through warehousing processes as the items were from an online clothing retailer. Therefore, further research into other warehouses and identifying different quality losses is advised.

Expanding this research into other warehouse areas is suggested and a final OWP using OEE model may be achieved after many different iterations and case studies. It is argued that a universal indicator of some sort using OEE can be achieved in the future, this can then be used for benchmarking alike warehouses.

Using weights for making certain coefficients more significant based on markup payments of contractors was done in this model. Including a cost factor in OEE coefficient can also be researched instead of using weights as it was criticized for negating interdependency of OEE coefficients.

Finally, quantitative analysis of the performance of OEE as an OWP compared to other used methods like DEA should be part of further research after validating the utility of a universal indicator.

Chapter 7: Conclusions

The aim of this research was to explore a new measure of overall warehouse performance (OWP). A literature review about the different warehouse types and activities was done. Warehouse KPIs were discussed and the need for an overall measure was established. Different methods of measuring OWP were briefly discussed, and it was identified that the most common measure of OWP in the literature was made using data envelopment analysis (DEA). The fundamental principles of production theory, such as returns to scale and production possibility set are used in DEA. That may be the root cause of DEA not improving performance assessment and benchmarking in non-production contexts like warehousing. These concepts are difficult to apply to pure multiple-criteria assessment problems, which are commonly solved using DEA.

A new method of measuring OWP was explored in a case study with an industry leader in logistics. The logistics service provider operates multiple warehouses owned by many clients. An overall metric that is widely accepted in the manufacturing industry was studied to explore its adaptability to a different industry, namely warehousing. The overall equipment effectiveness (OEE) metric has been utilized for many years in the production and manufacturing field and has been favored by many managers. Managers value this indicator because it is an overall metric that is simple and straightforward in identifying areas for potential improvement. OEE evolution was noted and the applicability and adaptation of OEE to other industries was identified, answering the second sub question for this project. This motivated to further explore the adaptability of OEE into warehousing and provide a modified OEE as a measure of OWP.

As OEE was initially developed for singular production equipment, it had to be modified to use the concept in warehousing. Much research was made on the evolution of OEE and case studies where OEE was modified to suit different particular industries are abundant. However, no research about the use of OEE in warehousing was made. Therefore, a challenge to identify the components of a modified OEE for warehousing was present. The original OEE concept measures the product of availability, performance, and quality of a singular equipment. Availability is a measure of the real time utilized during a shift. Performance is concerned with the speed of this production. It compares the actual output to the nominal output levels based on the available time. And quality measures the percentage of conforming products made by this equipment.

This thesis was done using a design science research (DSR) approach to achieve the objectives of this research that align with the ultimate goal of DSR in solving a real-life problem. The main objective was to develop a framework to measure the OWP of an automated warehouse using a modified OEE. Firstly, the original OEE was modified by identifying the major warehouse losses based on input from experts in the warehousing industry. Multiple interviews and detailed warehouse tours took place to draw out important KPIs and critical success factors (CSFs) that an overall metric should consider. Dock to stock times, labour productivity, and storage accuracy were marked as the most important. This answers sub question 1 of this research. A framework for a modified OEE for warehouses was developed using these inputs. Additionally, a modified OEE formula for the automated warehouse was made based on the major losses that occur in such warehouse. This framework was the basis for creating a software artefact to calculate OEE and its coefficients as well as coefficient components. An artefact was developed and the steps of creating it were presented. The artefact was evaluated by experts through interviews. Evaluation was based on multiple criteria, namely the complexity of the artefact, applicability and relevance of the model, and the applications intended for the use of OEE in warehousing.

Many uses of OEE have been mentioned in the literature and in practice. Particular uses of the model presented in this research were stated and evaluated. It was agreed that one of the main uses of an OWP measure lies in reporting performance to different stakeholders, mainly the warehouse contractor. Since the model is made from a service provider's perspective, effectiveness of their operations can be showcased using the modified OEE model. Using the model provides some more detail to discuss coefficients separately. This brings upon another major use of OEE within continuous improvement. Areas of improvement are highlighted by the model. And the impact of downtimes on performance can be identified by the speed loss value the model calculates, answering the third sub question of this research. The impact of certain changes and developments on the overall measure can be observed to identify the usefulness of that change.

Drawing OEE trends over time to identify developments in company performance can be an easy and straightforward way to show the effectiveness of management strategy and operational excellence. Tracking OEE day by day or shift by shift is possible. This tracking amongst the previously noted areas of continuous improvement show how the model can be used a science-based ground for improvement, satisfying the final research sub question. Using the measure as a benchmark for warehouses is difficult at the current stage of the model. This is due to the different definitions of losses amongst various warehouse types. The nature of the items processed at a warehouse and the warehouse utilization stage also impacts the KPIs to be considered by a modified OEE model.

By definition, OEE coefficients are independent of each other. However, upon studying the system, a hidden impact of downtimes on performance was noted and theoretical implications were made. Practical implications were discovered, and explanations were suggested. The model highlights the speed loss in performance which may resemble some of this hidden impact. Reasons for this effect were identified. Tradeoffs between OEE coefficients may exist and are avoided by setting high minimum thresholds for certain coefficients.

The research was done using a single case study in the inbound part of an automated warehouse for online clothing retail. There are many more types of items and warehouses that may require different definitions of warehouse losses and OEE coefficients. Expanding this research into other warehouse areas is suggested and a final OWP using a modified OEE model may be achieved after numerous iterations and case studies. It is argued that a universal indicator of some sort using OEE can be achieved in the future, this can then be used for benchmarking alike warehouses. Finally, quantitative analysis of the performance of OEE as an OWP measure compared to other used methods like DEA should be part of further research after validating the utility of a universal indicator.

References

Bamber, C., Castka, P., Sharp, J. M., & Motara, Y. (2003). Cross-functional team working for overall equipment effectiveness (OEE). *Journal of Quality in Maintenance Engineering*, *9*, 223–238. https://doi.org/10.1108/13552510310493684

Brocke, J. vom, Hevner, A., & Maedche, A. (2020). *Introduction to Design Science Research* (pp. 1–13). https://doi.org/10.1007/978-3-030-46781-4_1

Chen, P.-S., Huang, C.-Y., Yu, C.-C., & Hung, C.-C. (2017). The examination of key performance indicators of warehouse operation systems based on detailed case studies. *Journal of Information and Optimization Sciences*, *38*(2), 367–389. https://doi.org/10.1080/02522667.2016.1224465

Chow, G., Heaver, T. D., & Henriksson, L. E. (1994). Logistics Performance. *International Journal of Physical Distribution & Logistics Management*, *24*(1), 17–28. https://doi.org/10.1108/09600039410055981

Collatto, D. C., Dresch, A., Lacerda, D. P., & Bentz, I. G. (2018). Is Action Design Research Indeed Necessary? Analysis and Synergies Between Action Research and Design Science Research. *Systemic Practice and Action Research*, *31*(3), 239–267. https://doi.org/10.1007/s11213-017-9424-9

Costa, A. F., Pinto, A., Mendes, A. C., Quinteiro, A. R., Coelho, C., Meira, C., Pereira, I., Marinho, J., Barbosa, J., Cunha, N., Miranda, R., Drumond, N., & Lima, R. M. (2023). *Key Performance Indicators: A Perspective on Developing and Implementing Performance Indicators in a Textile Company Warehouse* (pp. 317–328). https://doi.org/10.1007/978-3-031-09382-1_28

Cramer-Flood, E. (2022). Global ecommerce forecast 2022. In *Insider Intelligence*. https://www.emarketer.com/content/global-ecommerce-forecast-2022

Dal, B., Tugwell, P., & Greatbanks, R. (2000a). Overall equipment effectiveness as a measure of operational improvement – A practical analysis. *International Journal of Operations & Production Management*, *20*(12), 1488–1502. https://doi.org/10.1108/01443570010355750

Dal, B., Tugwell, P., & Greatbanks, R. (2000b). Overall equipment effectiveness as a measure of operational improvement – A practical analysis. *International Journal of Operations & Production Management*, *20*, 1488–1502. https://doi.org/10.1108/01443570010355750

de Groote, P. (1995). Maintenance performance analysis: a practical approach. *Journal of Quality in Maintenance Engineering*, *1*(2), 4–24. https://doi.org/10.1108/13552519510089556

Denning, P. J. (1997). A new social contract for research. *Communications of the ACM*, *40*(2), 132–134. https://doi.org/10.1145/253671.253755

Domingo, R., & Aguado, S. (2015). Overall Environmental Equipment Effectiveness as a Metric of a Lean and Green Manufacturing System. *Sustainability*, *7*(7), 9031–9047. https://doi.org/10.3390/su7079031

Dresch, A., Lacerda, D. P., & Antunes Jr, J. A. V. (2015). *Design Science Research*. Springer International Publishing. https://doi.org/10.1007/978-3-319-07374-3

Ericsson, J. (1997). *Disruption Analysis - An Important Tool in Lean Production*.

Faber, N., de Koster, R. B. M., & Smidts, A. (2018). Survival of the fittest: the impact of fit between warehouse management structure and warehouse context on warehouse performance. *International Journal of Production Research*, *56*(1–2), 120–139. https://doi.org/10.1080/00207543.2017.1395489

Faveto, A., Traini, E., Bruno, G., & Lombardi, F. (2021). Development of a key performance indicator framework for automated warehouse systems. *IFAC-PapersOnLine*, *54*(1), 116–121. https://doi.org/10.1016/j.ifacol.2021.08.013

Frazelle, E. H. (2002). *World-Class Warehousing and Material Handling*.

García-Arca, J., Prado-Prado, J. C., & Fernández-González, A. J. (2018). Integrating KPIs for improving efficiency in road transport. *International Journal of Physical Distribution & Logistics Management*, *48*(9), 931–951. https://doi.org/10.1108/IJPDLM-05-2017-0199

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355. http://www.jstor.org/stable/43825912

Hevner, March, Park, & Ram. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75. https://doi.org/10.2307/25148625

Huang, S. H., Dismukes, J. P., Shi, J., Su, Q., Wang, G., Razzak, M. A., & Robinson, D. E. (2002). Manufacturing system modeling for productivity improvement. *Journal of Manufacturing Systems*, *21*(4), 249–259. https://doi.org/10.1016/S0278-6125(02)80165-0

Iivari, J., & Venable, J. (2009). Action research and design science research - Seemingly similar but decisively dissimilar. In *17th European Conference on Information Systems, ECIS 2009*.

Islam, M. R., Ali, S. M., Fathollahi-Fard, A. M., & Kabir, G. (2021). A novel particle swarm optimization-based grey model for the prediction of warehouse performance. *Journal of Computational Design and Engineering*, *8*(2), 705–727. https://doi.org/10.1093/jcde/qwab009

Johannesson, P., & Perjons, E. (2014). *An Introduction to Design Science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-10632-8

Johnson, A., & McGinnis, L. (2011). Performance measurement in the warehousing industry. *IIE Transactions (Institute of Industrial Engineers)*, *43*(3), 220–230. https://doi.org/10.1080/0740817X.2010.491497

Jonsson, P., & Lesshammar, M. (1999). Evaluation and improvement of manufacturing performance measurement systems - The role of OEE. In *International Journal of Operations and Production Management* (Vol. 19, Issue 1, pp. 55–78). https://doi.org/10.1108/01443579910244223

Karim, N. H., Abdul Rahman, N. S. F., Md Hanafiah, R., Abdul Hamid, S., Ismail, A., Abd Kader, A. S., & Muda, M. S. (2021). Revising the warehouse productivity measurement indicators: ratio-based benchmark. *Maritime Business Review*, *6*(1), 49–71. https://doi.org/10.1108/MABR-03-2020-0018

Karlsson, C. (2016). *Research methods for operations management* (Vol. 2). Routledge London.

Kusrini, E., Novendri, F., & Helia, V. N. (2018). Determining key performance indicators for warehouse performance measurement - A case study in construction materials warehouse. *MATEC Web of Conferences*, *154*. https://doi.org/10.1051/matecconf/201815401058

Larrañaga Lesaca, J. M., Zulueta Guerrero, E., Lopez-Guede, J. M., Ramos-Hernanz, J., Larrañaga Juaristi, A., & Akizu, O. (2017). Measuring global effectiveness of integrated electric energy systems. *International Journal of Hydrogen Energy*, *42*(28), 18121–18133. https://doi.org/10.1016/j.ijhydene.2017.04.057

Lu, C. S., & Yang, C. C. (2010). Logistics service capabilities and firm performance of international distribution center operators. *Service Industries Journal*, *30*(2), 281–298. https://doi.org/10.1080/02642060802123392

Manikas, I., & Terry, L. A. (2009). A case study assessment of the operational performance of a multiple fresh produce distribution centre in the UK. *British Food Journal*, *111*(5), 421–435. https://doi.org/10.1108/00070700910957276

Moonen, H., Popova, V., & Krauth, E. (2005). *Performance Measurement and Control in Logistics Service Providing. PERFORMANCE MEASUREMENT AND CONTROL IN LOGISTICS SERVICE PROVIDING*. https://www.researchgate.net/publication/220709920

Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): Literature review and practical application discussion. *International Journal of Production Research*, *46*(13), 3517–3535. https://doi.org/10.1080/00207540601142645

Muñoz-Villamizar, A., Santos, J., Montoya-Torres, J. R., & Jaca, C. (2018). Using OEE to evaluate the effectiveness of urban freight transportation systems: A case study. *International Journal of Production Economics*, *197*, 232–242. https://doi.org/10.1016/j.ijpe.2018.01.011

Muthiah, K. M. N., & Huang, S. H. (2007). Overall throughput effectiveness (OTE) metric for factory-level performance monitoring and bottleneck detection. *International Journal of Production Research*, *45*(20), 4753–4769. https://doi.org/10.1080/00207540600786731

Nachiappan, R. M., & Anantharaman, N. (2006). Evaluation of overall line effectiveness (OLE) in a continuous product line manufacturing system. *Journal of Manufacturing Technology Management*, *17*(7), 987–1008. https://doi.org/10.1108/17410380610688278

Nakajima, S. (1988). *Introduction to TPM: total productive maintenance*. Productivity Press.

Nakhla, M. (2018). Designing extended overall equipment effectiveness: application in healthcare operations. *International Journal of Management Science and Engineering Management*, 1–10. https://doi.org/10.1080/17509653.2017.1373377

Ng Corrales, L. del C., Lambán, M. P., Hernandez Korner, M. E., & Royo, J. (2020). Overall Equipment Effectiveness: Systematic Literature Review and Overview of Different Approaches. *Applied Sciences*, *10*(18), 6469. https://doi.org/10.3390/app10186469

Ng Corrales, L. del C., Lambán, M. P., Morella, P., Royo, J., Sánchez Catalán, J. C., & Hernandez Korner, M. E. (2022). Developing and Implementing a Lean Performance Indicator: Overall Process Effectiveness to Measure the Effectiveness in an Operation Process. *Machines*, *10*(2), 133. https://doi.org/10.3390/machines10020133

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pintelon, L., Gelders, L., & van Puyvelde, F. (2000). *Maintenance management*. Acco.

Pinto, M. M. O., Goldberg, D. J. K., & Cardoso, J. S. L. (2017). Benchmarking operational efficiency of port terminals using the OEE indicator. *Maritime Economics & Logistics*, *19*(3), 504–517. https://doi.org/10.1057/mel.2016.6

Puvanasvaran, A. P., Yoong, S. S., & Tay, C. C. (2017). Effect of hidden wastes in overall equipment effectiveness calculation. *ARPN J. Eng. Appl. Sci*, *12*, 6443–6448.

Raouf, A. (1994). Improving Capital Productivity through Maintenance. *International Journal of Operations & Production Management*, *14*(7), 44–52. https://doi.org/10.1108/01443579410062167

Ratnasari, D., Nazir, F., Toresano, L. O. H. Z., Pawiro, S. A., & Soejoko, D. S. (2016). The correlation between effective renal plasma flow (ERPF) and glomerular filtration rate (GFR) with renal scintigraphy ^{99m}Tc-DTPA study. *Journal of Physics: Conference Series*, *694*, 012062. https://doi.org/10.1088/1742-6596/694/1/012062

Rogers, P. (2021). *Glossary of Procurement and Supply Chain Terms*. CIPS. https://www.cips.org/intelligence-hub/glossary-of-terms

Sandberg, E., Oghazi, P., Chirumalla, K., & Patel, P. C. (2022). Interactive research framework in logistics and supply chain management: Bridging the academic research and practitioner gap. *Technological Forecasting and Social Change*, *178*, 121563. https://doi.org/10.1016/j.techfore.2022.121563

Scott, D., & Pisa, R. (1998). Can overall factory effectiveness prolong Mooer's law? *Solid State Technology*, *41*(3), 75+. https://link.gale.com/apps/doc/A21188455/AONE?u=googlescholar&sid=bookmark-AONE&xid=6461ee2b

Shani, A. B. (Rami), & Coghlan, D. (2019). Action research in business and management: A reflective review. *Action Research*, *19*(3), 518–541. https://doi.org/10.1177/1476750319852147

Staudt, F. H., Alpan, G., di Mascolo, M., & Rodriguez, C. M. T. (2015). Warehouse performance measurement: A literature review. In *International Journal of Production Research* (Vol. 53, Issue 18, pp. 5524–5544). Taylor and Francis Ltd. https://doi.org/10.1080/00207543.2015.1030466

ten Hompel, M., & Schmidt, T. (2007). *Warehouse management: automation and organisation of warehouse and order picking systems*. Springer.

Tona, K.-D., van Osch, M. J. P., Nieuwenhuis, S., & Keuken, M. C. (2019). Quantifying the contrast of the human locus coeruleus in vivo at 7 Tesla MRI. *PLOS ONE*, *14*(2), e0209842. https://doi.org/10.1371/journal.pone.0209842

Torabizadeh, M., Yusof, N. M., Ma'aram, A., & Shaharoun, A. M. (2020). Identifying sustainable warehouse management system indicators and proposing new weighting method. *Journal of Cleaner Production*, *248*, 119190. https://doi.org/10.1016/j.jclepro.2019.119190

Ulmer, M. (2017). Delivery deadlines in same-day delivery. *Logistics Research*, *10*(1). https://doi.org/10.23773/2017_3

van den Berg, J. P., & Zijm, W. H. M. (1999). Models for warehouse management: Classification and examples. In *Int. J. Production Economics* (Vol. 59).

Venable, J. (2009). Identifying and Addressing Stakeholder Interests in Design Science Research: An Analysis Using Critical Systems Heuristics. In *CreativeSME 2009. IFIP Advances in Information and Communication Technology* (Vol. 301, pp. 93–112). https://doi.org/10.1007/978-3-642-02388-0_7

Warehousing and Storage Market: Global Industry Trends, O. (2022). *Warehousing and Storage Market Trends And Forecast 2022-2027*. Imarcgroup.com. Retrieved 10 August 2022, from https://www.imarcgroup.com/warehousing-and-storage-market.

Williamson, R. M. (2006). *Using Overall Equipment Effectiveness: the Metric and the Measures*.

Wojcik, V., Dyckhoff, H., & Clermont, M. (2019). Is data envelopment analysis a suitable tool for performance measurement and benchmarking in non-production contexts? *Business Research*, *12*(2), 559–595. https://doi.org/10.1007/s40685-018-0077-z

Wu, Y., & Dong, M. (2008). Combining multi-class queueing networks and inventory models for performance analysis of multi-product manufacturing logistics chains. *International Journal* *of Advanced Manufacturing Technology*, *37*(5–6), 564–575. https://doi.org/10.1007/s00170-007-1004-1

Zhang, Y., & Wildemuth, B. M. (2009). Unstructured interviews. *Applications of Social Research Methods to Questions in Information and Library Science*, 222–231.

Zowghi, D., & Coulin, C. (2005). Requirements elicitation: A survey of techniques, approaches, and tools. In *Engineering and managing software requirements* (pp. 19–46). Springer.

Appendices

Appendix A

Questions asked include the description of warehouse processes, the flow of items within the warehouse, general understanding of the automation levels, the responsibilities of each the three stakeholders (the client and its WMS, mechanization partner, and logistics service provider), the level of interaction between different stakeholders, some common issues that occur often, and the nature of recorded data and involved parties.

Appendix B

1. Introduction:

- What is your job at the company?
- Do you have any previous experience or knowledge with warehouse KPIs?
- Are you familiar with the term "overall warehouse performance"?
- Do you agree that I record the following part of our interview for conducting my thesis research? Only I will have access to the collected data and all recordings/answers are treated anonymously and confidentially.

2. Overall Perspectives on warehouse KPIs:

- What are the KPIs that are more influential on the warehouse performance?
- What are target KPIs or SLAs that need to be met to get paid by the client?
- How is labour productivity measured?
- Is labour productivity important in an automated warehouse?
- Are downtimes recorded?
- What quality measures are in place for inbound?
- How is quality maintained?
- How is the overall performance of the warehouse evaluated?
- How do stakeholders access performance data?
- How are errors reported?
- In your opinion, do some errors stand out more than others in terms of frequency or impact?
- Would you be able to recall all warehouse KPIs?

3. Specific OWP and OEE related questions:

- What are the requirements for an OWP measure?
- Are your interests as a logistics service provider different than that of the client?
- How often is your performance compared to other warehouses within CEVA?
- What factors do you believe define the success of a service provider like CEVA?
- Do you track your performance over time?
- Mention some concerns that you believe need addressing to increase effectiveness?
- Are you familiar with OEE?
- Is OEE calculated at your warehouse?
- Will an overall metric be useful for your company?

4. Closing questions and concerns:

- Do you think a modified OEE as a measure of OWP can be used in your company?
- If yes, name the most important use of it in your opinion?
- Would you like to add anything that was not covered in this interview?

Appendix C

1. Introduction:

- What is your job at the company?
- Do you have any previous experience or knowledge with Microsoft Excel?
- Are you familiar with the research that I was conducting?
- Do you agree that I record the following part of our interview for conducting my thesis research? Only I will have access to the collected data and all recordings/answers are treated anonymously and confidentially.

2. Overall Structure of the Artefact:

- What are your first impressions on the model presented to you?
- How complex is it for you?
- Do you believe this is applicable in practice?
- Would you think it is difficult to use this model?
- Are the fields clear to you?
- Are you familiar with all the terminologies?
- What is your overall impression based on the structure of the model?

3. Detailed Questions about the Artefact:

- Are the inputs of this model measured at your company?
- Are the fields to be filled related to each other?
- Mention any confusions you find towards this artefact?
- Are the three coefficients clear to you?
- Do you understand what is being measured by the model?
- Is this representative of the initial thoughts you had about an overall measure of your company's performance?
- What are your suggestions for the values of the weights?
- Do you think prioritizing one coefficient over the other is useful?
- Describe your satisfaction with this artefact?
- How useful do you think this artefact is?
- Can you name some uses where your company can utilize this artefact?

4. Closing Questions and Suggestions:

- What improvements can be made to this artefact?
- What is the most important one in your opinion if there are any suggestions?
- Are there any uncovered issues you would like to discuss?

Appendix D

This appendix is strictly for assessment purposes and shall only be viewed by the graduation committee. This appendix is to be removed from the published version of the thesis.