Aalto University
School of Science

Master's Programme in Industrial Engineering and Management

Improving performance metrics in factory production and intralogistics

A bottom-up case study of a Finnish forestry machinery manufacturer

---

Sami Pelander

Master's Thesis
2025

Aalto University
School of Science

Author Sami Pelander

Title Improving performance metrics in factory production and intralogistics — A bottom-up case study of a Finnish forestry machinery manufacturer

Degree programme Industrial Engineering and Management

Major Strategy

Supervisor Prof. Kari Tanskanen

Advisor M.Sc. Janne Kilpua

Collaborative partner Ponsse Plc

Date November 23, 2025 **Number of pages** 87+3 **Language** English

Abstract

While many industrial companies use metrics to quantify operational performance, they can struggle to understand the connection between the measured phenomena and organizational outcomes. This thesis contributes to the general understanding of performance measurement in discrete manufacturing by examining how strategic and operational factors jointly shape factory-level metrics, and by proposing a structured framework for aligning measurement practices with decision making needs.

Conducted as part of the TwinFlow research project, the thesis analyses the performance measurement practices of Ponsse Plc, a Finnish forestry equipment manufacturer, through a single case study supported by triangulation interviews with two additional original equipment manufacturers. Based on the literature review, a set of strategic and operational factors influencing metric formation was identified and applied to evaluate the case company's existing measurement system.

By analyzing improvement opportunities in the case company's measurement system and benchmarking best practices from comparable industrial firms, the research identifies three key development directions for Ponsse's measurement practices. These directions include enhancements to existing metrics used by the company while also introducing new measurement opportunities.

Building on a synthesis of academic literature and empirical evidence, the study develops a conceptual framework that links key performance dimensions (productivity, quality, flexibility, time, and cost) to both operational and strategic objectives. The research extends existing theory by demonstrating how some performance dimensions are particularly well-suited to supporting different organizational processes, enabling a more nuanced understanding of the roles of different performance metrics. Through this, the thesis provides generalizable insights for designing performance measurement ecosystems and offers practical guidance for organizations seeking to align their metrics with decision making needs across different managerial levels.

Keywords performance measurement , discrete manufacturing , productivity , metrics , continuous improvement , production , case study

Aalto-yliopisto
Perustieteiden
korkeakoulu

Tekijä Sami Pelander

Työn nimi Suorituskykymittarien parantaminen tehdastuotannossa ja sisälogistiikassa — "Alhaalta ylös" etenevä tapaustutkimus suomalaisesta metsäkonevalmistajasta

Koulutusohjelma Tuotantotalouden maisteriohjelma

Pääaine Strategia

Työn valvoja Prof. Kari Tanskanen

Työn ohjaaja DI Janne Kilpua

Yhteistyötaho Ponsse Oyj

Päivämäärä 23.11.2025 **Sivumäärä** 87+3 **Kieli** englanti

Tiivistelmä

Monet teollisuusyritykset hyödyntävät mittareita operatiivisen suorituskykynsä arvioinnissa, mutta mitattujen ilmiöiden ja organisaation kilpailukyvyn yhteyden ymmärtäminen voi olla käytännössä haastavaa. Tämä diplomityö syventää ymmärrystä suorityskyvyn mittaamisesta yksikkötuotannossa tarkastelemalla, kuinka strategiset ja operatiiviset tekijät vaikuttavat tehdastason mittareihin. Lisäksi työ esittää viitekehyksen mittaamiskäytäntöjen ja organisaation päätöksenteon tarpeiden yhteensovittamiseksi.

Työ on toteutettu osana TwinFlow-tutkimushanketta ja se analysoi suomalaisen metsäkonevalmistaja Ponsse Oyj:n mittarointikäytäntöjä tapaustutkimuksen ja triangulaatiohaastatteluiden kautta. Kirjallisuuskatsauksen perusteella tunnistettuja metriikkoihin vaikuttavia strategisia ja operatiivisia tekijöitä hyödynnetään tapausyrityksen nykyisen mittausjärjestelmän arvioinnissa.

Analysoimalla mittausjärjestelmän kehityskohteita ja vertaamalla käytäntöjä muihin teollisuusyrityksiin, tutkimus tunnistaa kolme keskeistä kehityssuuntaa Ponssen mittauskäytännöille. Nämä suositukset sisältävät parannuksia olemassa oleviin mittareihin sekä uusia mittausmahdollisuuksia tunnistettujen tarpeiden perusteella.

Kirjallisuuden ja empiirisen aineiston synteesin pohjalta työssä kehitetään käsitteellinen viitekehys, joka yhdistää keskeiset suorituskyvyn ulottuvuudet (tuottavuus, laatu, joustavuus, aika ja kustannukset) teollisuusyrityksen operatiivisiin ja strategisiin tavoitteisiin. Tutkimus syventää teoriaa osoittamalla, miten tietyt suorituskyvyn ulottuvuudet soveltuvat erityisen hyvin erilaisten organisaatioprosessien mittarointiin. Näin diplomityö tarjoaa yleistettäviä näkökulmia suorituskyvyn mittausjärjestelmien suunnitteluun sekä käytännön ohjeistusta organisaatioille, jotka pyrkivät sovittamaan mittaristonsa paremmin yhteen erilaisten päätöksentekotarpeiden kanssa.

Avainsanat suorituskyvyn mittaaminen, yksikkötuotanto, tuottavuus, metriikka, jatkuva parantaminen, tuotanto, tapaustutkimus

Preface and acknowledgments

This thesis marks the conclusion of my 5.5 years of studies at Aalto University. At least in my opinion, the final six months spent working on this thesis have been a fitting way to complete my degree. This study has allowed me to apply what I have learned in my IEM studies while also gaining a deeper understanding of the Finnish manufacturing industry.

When I began my studies in the middle of the COVID pandemic, it was difficult to foresee how meaningful both my studies and the everyday life around them would eventually become. Looking back, these university years have passed surprisingly quickly, and I have come to appreciate how easily student life can carry you along. I am grateful to Prodeko for enabling an active student life and fostering a spirit of inspiration and encouragement, to my classmates for making both course projects and the student life outside them genuinely enjoyable, and to the associations of Otaniemi for offering me hobbies, friends, responsibilities, and some of the most memorable experiences of my life.

I would like to express my sincere gratitude to my supervisor, Kari Tanskanen, and my instructor, Janne Kilpua, for their guidance, encouragement, and open communication throughout this process. I am equally thankful to the people at Ponsse for participating in the study, for their openness and commitment to advancing the research, and for the exceptional hospitality shown to us in Vieremä. It is easy to understand why the organization is consistently ranked among Finland's most reputable companies.

I also wish to thank my family for their continuous support throughout my studies. Finally, my deepest gratitude goes to my partner, Emma, for her unwavering encouragement and patience throughout the past eight years.

It is now time to take the next steps toward new challenges in working life. I am glad to know that while my studies will be over, I will still carry with me the friendships, experiences, and lessons that made these years so meaningful. It is also comforting to know that I am not quite ready to leave Otaniemi behind just yet, and that Teekkarikylä will remain part of my life through hobbies and leisure activities for the time being.

Helsinki, November 23rd 2025

Sami Pelander

Table of Contents

Abbreviations

OEM	Original Equipment Manufacturing
IL	Internal Logistics
ERP	Enterprice Resource Planning
MES	Manufacturing Execution System
ERP	Enterprice Resource Planning
CPS	Cyber-Physical Systems
IIoT	Industrial Internet of Things
WIP	Work in Process
TH	Throughput
CT	Cycle Time
CI	Continuous Improvement
TPS	Toyota Production System
JIT	Just-in-Time
5S	Sort, Set in order, Shine, Standardize, Sustain
KPIs	Key Performance Indicators
ABC	Activity-Based Costing
VSC	Value Stream Costing
OEE	Overall Equipment Effectiveness
TPM	Total Productive Maintenance
OFE	Overall Factory Effectiveness
OTE	Overall Throughput Effectiveness
MTO	Made-to-order
NPD	New Product Development
POC	Proof-of-Concept

List of Figures

List of Tables

1 Introduction

1.1 Research Background and Motivation

In the context of industrial manufacturing, performance measurement is a vital aspect of both operational and strategic management, as it enables the monitoring, evaluation, and improvement of recurring processes. At its core, performance measurement aims to make complex systems more comprehensible by quantifying actions into indicators that reflect the performance of interconnected tasks, workflows, and human actions. The applications for these indicators are wide-ranging, extending from day-to-day operational management to long-term strategic planning.

Modern management philosophies, such as Lean Management and Total Quality Management, also emphasize systematic measurement of performance as a foundation for organizational decision making. As noted by Neely et al. (1995), within these management frameworks, performance metrics should be not only tools to track output, but also mechanisms to identify discrepancies and align daily operations with strategic goals.

Despite its long-standing relevance, performance measurement remains a fragmented and evolving field. Literature on the topic spans from managerial philosophies to highly granular shop-floor metrics, encompassing multiple schools of thought. Although various frameworks and metrics have been proposed over the decades, there is no universal agreement on how performance should be measured or how metrics should be structured to support improvement in practice. This is particularly evident at the factory level, where dynamic environments, fragmented data systems, and complex workflows often hinder the effective application of performance metrics.

This thesis seeks to contribute to this area by analyzing how the multifaceted concept of performance is understood and monitored through metrics in discrete manufacturing. The objective is to identify ways to improve performance measurement so that it captures the most meaningful aspects of production and provides information that supports practical decision making across both operational and strategic levels.

The study follows a case study methodology in line with the the depictions of Eisenhardt and Graebner (2007) and Yin (2014), focusing on a Finnish case company operating in the forestry machinery sector. More specifically, this thesis conducts case research as a theory elaboration (Ketokivi and Choi, 2014), in which the empirical context is interpreted through the lens of an existing theoretical framework.

Building on a synthesis of academic research, this thesis proposes a framework for evaluating performance metrics in manufacturing processes. The framework is then applied to the practices of the case company to identify gaps between current measurement practices and the operational realities of factory-level production. Based on this evaluation, this thesis suggests context-sensitive metrics that better support operational decision making and strategic improvement initiatives. In addition, the results are used to elaborate on the theoretical framework to better understand how metrics should support organizational processes. While the findings are grounded in a single case, the objective is to generate analytical insights that may be transferable to other discrete manufacturing contexts facing similar challenges.

1.2 Research Context

This Master's thesis is written as a part of the *TwinFlow* project, an Aalto University -led research project involving multiple industrial manufacturing and software companies. The project is a part of Business Finland's Data Economy program, which aims to enhance the operative competitiveness of the Finnish manufacturing industry. The project consortium consist of 11 organizations, which can be divided into four categories according to their area of expertise:

- three large-size original equipment manufacturing (OEM) companies,
- three companies producing production machinery and related services,
- three companies providing technology and service solutions for industrial context, focusing e.g. on data-analytics and machine vision,
- two universities providing research teams with relevant knowledge on project topics, facilitating communication and providing testing facilities for solutions developed during the project.

The empirical research context for this thesis consists of the manufacturing operations of Ponsse Plc, one of the OEM companies participating in the project. Ponsse is a Finnish company founded in 1970, manufacturing a range of forestry machinery such as harvesters and forwarders (Ponsse, n.d.). The company currently manufactures all of their products in a single manufacturing plant located in Vieremä, Finland.

This thesis ties in with the task 2.1 of the *TwinFlow* project's consortium plan, focusing on the convergence of information technology and operational technology. More specifically, in addition to theory elaboration, the objective of this thesis is to review the perceived gaps in current performance metrics used by the case company in their production plant and to suggest and validate new improvement opportunities that have been identified through an on-site data collection phase and a review of relevant academic literature. Through this methodology, this thesis aims to provide insights into how the quantitative and qualitative data collected from the operations of Ponsse can be leveraged to provide higher-level monitoring and development opportunities for the company, in addition to identifying generalizable findings that can benefit other consortium partners of the *TwinFlow* project.

The scope of this thesis is limited to the production processes and internal logistics (IL) of Ponsse's Vieremä manufacturing plant. In the context of this thesis, IL will be used to refer to the material flows occurring between the inbound and outbound logistics of the selected manufacturing plants.

1.3 Research Questions

This thesis is structured around an overarching research question, which is supported by three sub-questions. The formulation of these questions is grounded in the academic discussion of performance measurement systems and their role in linking operational outcomes with broader organizational objectives. Prior research highlights that while performance metrics are essential for managing manufacturing operations, they often fail to capture the multidimensional and evolving nature of performance. Furthermore, recent developments in data availability and digital manufacturing have emphasized the need to reassess how metrics are defined, implemented, and interpreted in industrial settings.

Against this background, the main research question seeks to enhance the understanding of the role of factory-level performance metrics in manufacturing and their cascading effects on organizational decision making and long-term productivity development:

RQ: *In what ways can factory-level performance metrics be refined to enable more effective decision making and contribute to sustained productivity improvement in discrete manufacturing environments?*

The main research question reflects the central aim of the study. In practice, answering the question involves analyzing the strengths and weaknesses of existing measurement practices through a literature review and a case study. The case study provides empirical data on existing practices, while the literature review provides the theoretical foundation against which the empirical findings are interpreted. In addition to studying the link between measurement and organizational performance, this research aims to propose concrete improvements to the practices of the case company by identifying blind spots in its current measurement system.

The overarching research question is further explored through the following three sub-questions:

SQ1: *What are the strengths and limitations of commonly used performance metrics in manufacturing?*

The first sub-question is addressed primarily through the literature review. It examines existing academic perspectives on performance measurement to identify how metrics contribute to organizational value creation and how they support continuous improvement processes.

SQ2: *What are the key improvement areas in the performance measurement practices of the case company, and how can metrics be aligned with them?*

The second sub-question is answered primarily through the empirical case study by identifying gaps and opportunities for improvement in the case company's measurement practices. This question also examines how the observed practices differ from the frameworks presented in the literature and aims to uncover the underlying reasons for these discrepancies.

SQ3: *What types of data and system capabilities are required to enable more accurate and relevant performance measurement?*

The third sub-question bridges the theoretical and empirical perspectives by analyzing the informational and technological requirements needed to implement more accurate and relevant performance metrics in practice. It aims to identify both enabling and constraining factors that affect the effectiveness of performance measurement systems in discrete manufacturing environments.

Together, these sub-questions structure the research process and provide a systematic approach to understanding the role of performance metrics in improving manufacturing performance. Answering these questions aims to offer both theoretical contributions in the form of theory elaboration and practical development implications for discrete manufacturing companies, with a particular focus on the case company examined in this study.

1.4 Structure of the thesis

This thesis is structured as follows:

- The second section presents a literature review on performance measurement in discrete manufacturing and presents a conceptual framework for aligning performance metrics with key strategic and operational factors.

- The third section describes the research methodology and design, as well as the data collection and analysis process.

- The fourth section presents the empirical results of the case study, including both the analysis of the case company current operations and the introduction of three targeted improvement opportunities.

- The fifth section discusses the research findings in relation to the theoretical background. It elaborates on the proposed performance measurement framework and outlines the limitations of the study and directions for future research.

2 Theoretical Background

This chapter presents a synthesis of existing academic knowledge on performance measurement and improvement approaches in discrete manufacturing, aimed at answering the previously presented research question and sub-questions while identifying research gaps in manufacturing performance measurement.

The first subchapter outlines the characteristics of discrete manufacturing processes, providing general context into the operational realities of the case company. The second subchapter discusses the concepts of productivity and performance in industrial setting, detailing common performance dimensions and their relevance in evaluating manufacturing operations. The third subchapter focuses on concrete approaches used to measure and improve productivity, covering both managerial frameworks and specific metrics commonly employed in practice. The fourth subchapter identifies both strategic and operational factors affecting performance measurement. Finally, the fifth subchapter draws conclusions from these findings and introduces a framework that offers guidance on how strategic goals and operational conditions affect the measurement practices of organizations.

2.1 Characteristics of Modern Discrete Manufacturing

Discrete manufacturing is generally described as the production of separate, distinct entities, such as automobiles or electronics (Helu et al., 2020). This is often contrasted with process manufacturing, where end products are created through a chemical transformation, such as the production of pharmaceuticals or fuels (Litster and Bogle, 2019). Both of these manufacturing processes generally consist of independent machines or workstations operating together to transform materials into end products. However, the nature of the end products significantly affects the structure and organization of the production process itself.

The two main elements that affect the structure of production systems are the *volume of output* and the *variability* of the end products. To illustrate this, figure 1 presents a product-process matrix introduced by Hayes and Wheelwright (1979), a widely recognized framework that relates product types to corresponding process structures. The matrix highlights fundamental relationships in manufacturing, displaying how manufacturers are required to align their production processes based on the products they manufacture.

Although the underlying logic of the product-process matrix remains valid to this day, production processes themselves are constantly adapting to external changes. Historically, discrete manufacturing processes have evolved from more static mass production lines, exemplified by the Tayloristic-Fordist model, towards a paradigm that increasingly emphasizes continuous improvement and operational adaptability (De Toni and Tonchia, 2002). This shift has been driven by larger changes in the global manufacturing environment. As global competition has increased, manufacturing companies have been required to comply with shorter delivery times, increased demand for product variety, and rapid adjustments for production capacity (Mehrabi et al., 2000).

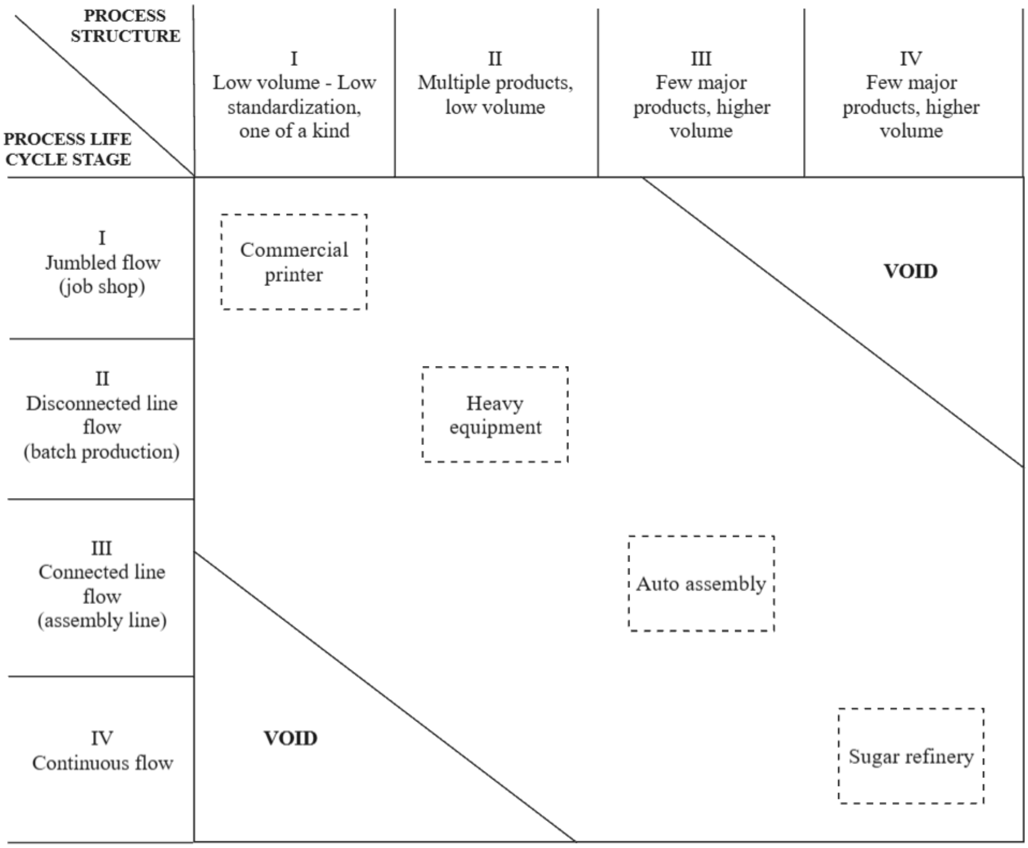

Figure 1: The product-process matrix as presented by Hayes and Wheelwright (1979).

These requirements have also influenced the inner workings of discrete manufacturing plants, resulting in more modular and distributed production systems (Helu et al., 2020). As a result, measuring the production capacity of modern discrete manufacturing systems has become challenging due to their increasingly complex nature. Daily production planning is affected by changes in customer needs, product types, and production factors, basing capacity prediction on different estimates instead of accurate and all-encompassing theoretical models (Cai et al., 2024).

Therefore, competitiveness in the manufacturing industry is largely based on the organizations' capability to evaluate the performance of their own operations and improve them accordingly. To enable this type of evaluation, scholars have formulated multiple different "performance factors" that are considered the most critical for measuring success (Wiendahl et al., 2015; Schroeder et al., 2002; Neely et al., 1995). Some common performance factors include:

- **Production costs:** A resource-oriented factor that accounts for monetary costs used for materials, personnel, and the production process itself.

- **Throughput time:** The total time required for an individual product to move through a production system.

- **Innovation capacity:** The ability of a manufacturing system to integrate new technologies, adapt to changing market demands, and develop new products or processes.
- **Flexibility:** The capability of a manufacturing system to adapt to changes in product type, volume, or production schedule with minimal loss in performance or efficiency.
- **Product and process quality:** The degree to which end products meet the design specifications and customer expectations, as well as the consistency and reliability of the production processes.

Although the relevance of these performance factors varies depending on the use case, they remain central to evaluating operational performance in most discrete manufacturing settings. Importantly, not all performance factors aim to be optimized in the same way. For example, while production costs and throughput time are generally minimized, factors such as product quality or flexibility require achieving and maintaining a sufficient level to meet customer expectations or strategic goals. They also differ in their applicability, as not all performance factors are easily or directly measurable. While factors such as production costs and throughput time can often be quantified with existing Enterprise Resource Planning (ERP) or Manufacturing Execution System (MES) systems (Xu et al., 2024), the innovation capacity or flexibility of an organization may need to be evaluated through indirect indicators. These may include metrics like *time-to-market for new products* or *changeover frequency* in the production process.

Another major element affecting the applicability of performance factors is the limited visibility into the manufacturing process itself. As Liukkonen and Tsai (2016) put it, the lack of real-time data acquisition and the limited traceability of production-related information results in low production visibility, hindering performance monitoring. This is seen as a common problem for companies operating in the manufacturing sector, where complex workflows are common. In many cases, this problem is alleviated through process simulation (e.g. Huang et al.,2003; Fabri et al.,2022). However, a more modern approach to combatting this phenomenon is linked to Industry 4.0, which aims to enable more real-time tracking of the production process.

2.1.1 Effects of Industry 4.0 on Discrete Manufacturing

Currently, one of the largest forces driving change in the global manufacturing industry is the Industry 4.0 revolution. Introduced in Germany in 2011 (Kagermann et al., 2013), Industry 4.0 refers to the integration of digital technologies into traditional manufacturing systems. The term covers a wide range of technologies aimed at dynamic data acquisition and analytics, such as Cyber-Physical Systems (CPS), Industrial Internet of Things (IIoT), big data analytics, collaborative robotics, additive manufacturing and augmented reality (Silvestri et al., 2022).

This digitalization of operations has major implications for productivity measurement and performance monitoring in discrete manufacturing. As Kamble et al. (2020) explain, Industry 4.0 enables the information flows of physical production processes to be available digitally in real time. A central enabler in this shift is the Industrial Internet of Things. In traditional discrete manufacturing systems, the lack of transparency and real-time data access often creates information silos, where individual machines or departments are operated without clear visibility of upstream or downstream processes. Yuan et al. (2013) describe this as the "phenomenon of information isolated islands". IIoT addresses this challenge by embedding sensors and connectivity into machines and products, allowing managers to track the movement and condition of parts throughout the production process.

In practice, this connectivity is often enabled by automatic identification technologies connected to physical products, such as magnetic recognition or optical/electronic tags. The operational data gathered through these technologies is then connected to the intranet of a manufacturing company, where it can then be utilized in e.g. decision making and further process automation (Liukkonen and Tsai, 2016). While not all companies have yet implemented these technologies, they represent a modern target state where organizations can more effectively measure the aforementioned performance factors.

Therefore, the adoption of Industry 4.0 technologies should help companies gain deeper insights into their operational processes. However, this information must then be utilized in ways that support both the operational and strategic goals of the organization. This consideration leads to a more detailed examination of how performance is measured in the context of industrial manufacturing, and how it differs from the concept of productivity.

2.2 Foundations of Performance in Manufacturing

2.2.1 The Concepts of Productivity and Performance

In its simplest form, productivity can be viewed as a straightforward indicator that describes the relationship between an output and the inputs that are required to generate that output (Schreyer and Pilat, 2001). In the context of discrete manufacturing, this type of productivity is used as a measure of the efficiency of the transformation process itself. By monitoring the monetary value of inputs used in the production process (e.g. purchase price of raw materials) and the outputs (e.g. sales price of end products), the total added value of the manufacturing process can be determined. Furthermore, these types of productivity measures are traditionally categorized either as *partial productivity measures* or *total productivity measures* (Misterek et al., 1992). Partial productivity measures represent the ratio of output to only one input, such as labor, while total productivity measures compare all outputs with all inputs. This traditional view of productivity is represented in figure 2.

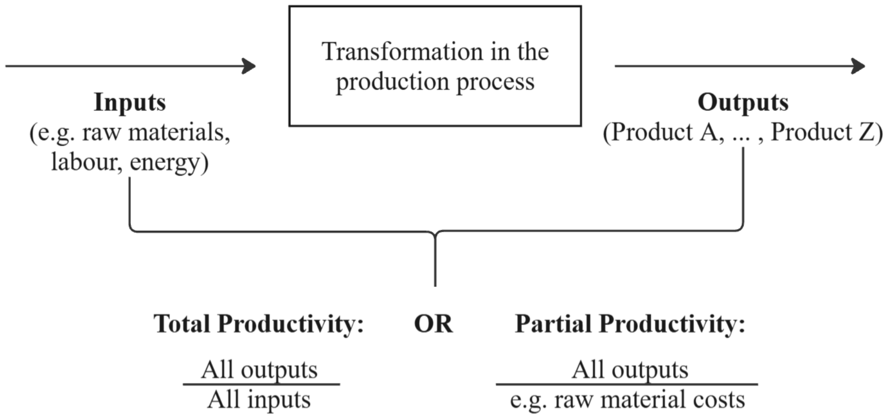

Figure 2: Productivity as Outputs/Inputs, adapted from (Misterek et al., 1992).

While total and partial productivity measures have long been used in industrial settings, their formalization and widespread use in management practices gained prominence during the 1970s (Hon, 2005). Most commonly, this type of productivity is measured using monetary values of goods and services, as they provide a uniform and objective basis for assessing the added value of the manufacturing process. Intuitively, this type of productivity can be maximized by increasing the throughput of a factory while minimizing the resources used in the manufacturing process. Despite their simplicity, these metrics can offer a useful starting point for identifying potential areas of improvement. Misterek et al. (1992) emphasize how one of the most important characteristics of productivity measures is their ability to reveal which specific cost elements contribute to changes in productivity. By comparing the value of outputs to individual variables such as labor or energy, a manufacturing company is able to better understand the root causes behind productivity growth or decline.

However, due to the simplicity of these measures, their practical applications in manufacturing are noticeably limited in their capabilities. As Misterek et al. (1992) also note, these measures often do not account for changes in the quality of inputs and outputs, and therefore do not encourage companies to evaluate the benefits associated with quality. In addition, the measures are highly sensitive to factors such as capacity utilization. For example, a company investing in a new manufacturing line may experience a temporary decline in productivity due to reduced capacity utilization despite the investment being strategically beneficial in the long term.

Due to these limitations, productivity measures alone are often not considered sufficient indicators of organizational performance or competitiveness. This recognition led to the emergence of multidimensional performance measures during the 1990s. In particular, the works of Eccles (1991) and Kaplan and Norton (1992) brought forth the idea that organizations should expand their focus beyond financial performance only. This multidimensional view of performance is based on the idea that an organization should comprehensively address the needs of different stakeholders. This includes taking into account factors such as customer satisfaction and the innovation capabilities of the organization in order to ensure competitiveness also in the future. From this perspective, productivity is reinterpreted as a subset of the broader concept of performance. Whereas productivity seeks to quantify efficiency in terms of resource use, performance encompasses additional strategic dimensions such as responsiveness, flexibility, product quality, and alignment with long-term organizational goals.

The multidimensional view of performance has also been explicitly applied in manufacturing contexts. Both Neely et al. (1995) and Hon (2005) note how manufacturing companies tend to evaluate their performance by following multiple operational metrics simultaneously. These metrics can usually be aggregated into categories that resemble the performance factors introduced in chapter 2.1. For example, a survey conducted by Hon (2005) identified a total of 442 individual metrics used by manufacturing companies, which were then categorized under five dimensions: *productivity*, *quality*, *flexibility*, *time* and *cost*. This multidimensional view of company performance is represented in figure 3.

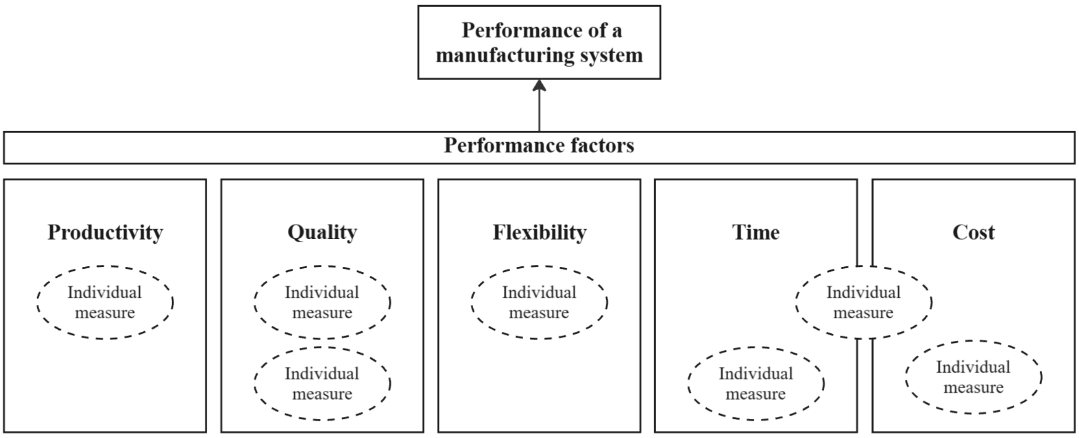

Figure 3: Example of performance measurement dimensions (factors) in a manufacturing system, adapted from Neely et al. (1995) and Hon (2005).

The multidimensional view has gained worldwide popularity after the 1990s, as it aims to patch the innate gaps of traditional productivity metrics. However, the relationships between different performance factors and the measures used often remain unclear. The early proponents of the multidimensional view proposed the idea of a "balanced scorecard", implying that a company should focus equally on all aspects of performance (Kaplan and Norton, 1992). In reality, however, the complexity and uncertainty of operating environments vary significantly between manufacturing companies, making the design of measurement systems highly context-dependent (Hon, 2005). Therefore, assigning even weight to all components of performance is often not practical or beneficial for company operations, making the balanced scorecard more of a tool for strategy communication.

Another common limitation in applying the multidimensional view is the lack of clarity regarding the number and scope of performance measures. This is also related to variance in operating environments, which requires each company to use alternative sets of measures (Singh et al., 2000). Adding redundant performance measures requires additional resources and can obscure the monitoring of the most relevant performance factors. As such, developing an effective performance measurement system requires balancing coverage and clarity to ensure that key operational dimensions are tracked while avoiding unnecessary complexity. The next subchapter discusses some of the most common measures used to gauge factory-level performance in the context of discrete manufacturing.

2.2.2 Factory-level Operations and Measurements

Monitoring performance at the factory level is critical for manufacturing companies, as this is where value is directly added through material transformation. Additionally, factory operations represent a contained and controllable environment in which processes can be directly affected by improvement activities. From a managerial point of view, the factory serves as the most concrete point of intervention, making it a natural focus for productivity and performance measurement and improvement efforts.

As noted by Grünberg (2003), efforts to improve productivity in manufacturing operations have existed in different forms since the beginning of the industrial era. Over time, these efforts have led to a wide variety of frameworks and tools aimed at optimizing factory-level processes. However, as Hopp and Spearman (2011) critically argue, there is still a lack of a scientific framework for manufacturing management. They mention how this fundamental problem results in multiple issues for factory-level performance measurement:

1. There is no universally accepted definition of the problem of improving productivity.
2. There is no uniform standard for evaluating competing policies.
3. There is little understanding of the relations between financial measures and manufacturing measures.
4. There is little understanding of how the manufacturing measures relate to each other.

These issues exemplify the fragmented nature of performance measurement practices in industrial manufacturing. Despite these limitations, there exist some fundamental relationships that are widely used to manage and analyze factory performance. One of the most fundamental and widely referenced "laws" in manufacturing is Little's law, which describes the relationship between three key variables in a stable production system:

$$\mathrm{WIP} = \mathrm{TH} \times \mathrm{CT} \tag{1}$$

Where *WIP* (work in process) represents the inventory between the start and end of the production process, *TH* (throughput) represents the average rate at which a manufacturing system produces finished products, and *CT* (cycle time) represents the total time to complete the production of a single unit from start to finish. Little's law provides a basis for scientific analysis of manufacturing and for evaluating fundamental trade-offs between inventory levels and processing speed.

Little's law also forms the theoretical foundation for *throughput analysis*, a method for evaluating and optimizing the throughput levels of systems. As described by Li et al. (2009), throughput analysis is an important tool for both the design and management of production systems, as it helps to assess the efficiency of a system and identify production bottlenecks. While conceptually simple, accurately modeling throughput is immensely difficult in real-world factories. As noted by both Li et al. (2009) and Hopp and Spearman (2011), exact throughput modeling typically requires the assumption of no variability, which is rarely valid in real-life discrete manufacturing systems. Consequently, throughput is more often estimated using historical performance data and observations than mathematical modeling.

The primary reason for this modeling difficulty is the presence of variability. As emphasized by Hopp and Spearman (2011), variability is the key disruptive factor in production environments, as it reduces the predictability of throughput. Variability in manufacturing can originate from multiple different sources, such as machine failures, material quality defects, order mix changes or skill differences between employees. In addition, it can also be introduced deliberately, for instance when a company expands its product portfolio.

The effects of variability are directly related to the broader performance factors discussed in Chapters 2.1 and 2.2.1, as increased variability often leads to lower productivity and higher costs. On the other hand, Ramasesh and Jayakumar (1991) mention how flexibility describes the ability of an manufacturing line to deal with variability. To cope with variability, Hopp and Spearman (2011) propose the concept of three fundamental buffers in production systems:

– **Inventory buffer:** Additional raw materials or work-in-progress inventories.

– **Capacity buffer:** Additional production resources, such as extra machines or labor.

– **Time buffer:** Slack in production schedules or longer lead times.

These buffers intend to reduce variability by minimizing the effects caused by disruptions in the manufacturing system. For example, because of inventory buffers, consecutive machines inside a factory can operate more independently of each other, reducing the disruptive downstream effects that e.g. machine breakdowns can cause. All three buffers serve as critical levers for stabilizing operations, but they also incur trade-offs, such as increased production costs. This creates the fundamental dynamic in factory-level production, where operations need to be optimized to maximize throughput while minimizing buffer levels.

Because there is no universally optimal way to manage variability (Hopp and Spearman, 2011), manufacturing companies must implement targeted improvement activities based on their operational context. These improvement efforts are typically guided by established methodologies or measurement systems. The following chapter explores these approaches in more detail.

2.3 Performance Improvement Approaches

Even though concepts such as Little's law provide the basis for relationships in the manufacturing process, improving performance involves much more than applying isolated formulas or theoretical models. In practice, factory-level performance is often improved through a variety of different methodologies that aim to translate performance-related goals into actionable frameworks that can be applied in day-to-day manufacturing operations. However, despite the widespread attention that the topic has received, there is no universally accepted methodology that guarantees performance improvement in all manufacturing contexts (Gershwin, 2000).

In their review of the literature on the topic, Muthiah and Huang (2006) classify manufacturing performance improvement efforts into four different categories:

– **Operations research -based methods:** Approaches that emphasize mathematical modeling as a tool in understanding the behavior of complex systems. Common methods include *linear/non-linear programming*, *simulation*, *graph theory*, and *queuing theory*.

– **System analysis -based methods:** The system analysis approach focuses on breaking down the manufacturing production system into its individual components and analyzing their interactions. Common methods include *Graphs with Results and Actions* and *Structured Analysis and Design Technique*.

– **Continuous improvement methods:** Methods, that are built around the central idea of eliminating waste in the manufacturing process. These methods are often closely tied to management philosophies, as they emphasize iterative problem-solving to drive long-term operational excellence. Common improvement methodologies include *Lean Manufacturing*, *Total Quality Management*, *5S*, and *Theory of Constraints*.

– **Performance metrics -based methods:** These methods are more focused on measuring the productivity and efficiency of systems. They provide mathematical metrics that are designed to evaluate the effectiveness of production. Common performance metrics include *Overall Equipment Effectiveness*, and *Overall Factory Effectiveness*

Even in this brief review of performance measurement approaches, the methods that Muthiah and Huang (2006) list are a mix of analytical, conceptual, and practical frameworks. Each can also address different levels of abstraction within the manufacturing system. This exemplifies the fact that performance as a concept can be analyzed from a multitude of different points of view. Therefore, instead of an universally applicable solution, the choice of measurement approach depends heavily on the organizational context, available resources and the intended use of the resulting insights.

For the purposes of this thesis, the focus will primarily be on productivity measurement approaches that align with the previously introduced **continuous improvement methods** and **performance metrics -based methods**. This is in part due to the scope of our research, as the objective is to formulate solutions that fit the needs of a single manufacturing facility. Another factor affecting this choice is the strong synergy between continuous improvement philosophies and the previously introduced Industry 4.0 technologies. As noted by Silvestri et al. (2022), the relationship between Industry 4.0 enabling technologies and the continuous improvement methods, such as Lean Production, has been extensively explored in academic literature. They further emphasize this dependency, stating that the concept of Industry 4.0 works only if it is integrated with the concepts of continuous improvement. By examining both common improvement methodologies and factory-level metrics, this thesis aims to identify measurement approaches that are both actionable for managers and grounded in actual production data, providing a practical basis for improving performance.

Furthermore, we exclude **operations research -based methods** and **system analysis -based methods**, as they tend to focus on mathematical optimization models or broader enterprise-level modeling, which typically require extensive proprietary data or broad generalizations. Moreover, implementing such methods would likely require prolonged data collection periods and resource commitments beyond the scope and timeframe of this thesis.

The following subchapters further discuss the two highlighted methods, their applicability to manufacturing context, and the concrete metrics associated with them.

2.3.1 Continuous Improvement Methods

In contrast to the previously discussed methods that focus on system modelling and quantitative analysis, Continuous improvement (CI) methods approach performance improvement from a more practical and action-oriented perspective. They encompass a wide range of management frameworks that emphasize the iterative nature of improvement, where incremental changes are systematically implemented, evaluated, and refined over time.

CI methods are considered to originate from the introduction of the Toyota Production System (TPS) led by Ohno (1988). TPS was formulated as a managerial philosophy that defines Toyota's method of manufacturing automobiles, and it focuses around the central idea of eliminating waste in the manufacturing process. This waste consists of any activity that does not add value from the customer's perspective, commonly categorized into seven types: defects, excess inventory, waiting, transport, overproduction, overprocessing and unnecessary movement (Ohno, 1988). This perspective on waste is used as a tool for systematically identifying non-optimal activities in the manufacturing process, providing a basis for iterative improvement efforts.

One of the key concepts behind TPS is *Just-in-Time* (JIT), which describes the ideal production system in which each process produces only what is needed by the next process, in the exact amount needed, and at the exact time it is needed (Ohno, 1988). The goal of JIT is to minimize inventory levels and the throughput time of a production system while maintaining a continuous flow of products. Combined with the goal of waste minimization, JIT enforces the idea of an optimized production system which aims to answer customer needs as well as possible.

As TPS and its core principles grew in popularity, they were gradually adopted under a broader managerial philosophy known as Lean Production. Notably, the work of Womack et al. (1990) played a central role in popularizing the term "Lean" and demonstrating its applicability outside of the automotive sector. Lean production and other similar CI methods have since become commonplace in production management worldwide. As noted by Jasti and Kodali (2015), these approaches are widely associated with positive organizational performance results. This is supported by a 2004 survey of 967 American manufacturing companies. which revealed that more than 95% of them had an established CI methodology in place (Taninecz, 2004).

To implement continuous improvement in practice, companies utilize a variety of standardized tools and techniques developed under the Lean framework. Some commonly used techniques include *Value Stream Mapping*, a framework that allows organizations to visualize material and information flows to identify and eliminate non-value-adding steps, and *5S* (Sort, Set in order, Shine, Standardize, Sustain), a guideline for workplace organization and cleanliness (Jasti and Kodali, 2015). Both of these techniques focus on identifying development targets within a company's current operations, after which different development actions are taken to eliminate inefficiencies.

The iterative nature of continuous improvement is often conceptualized through models such as Deming's wheel of improvement (Grünberg, 2003), which emphasizes feedback gathering as an important factor of CI. This logic is illustrated in figure 4:

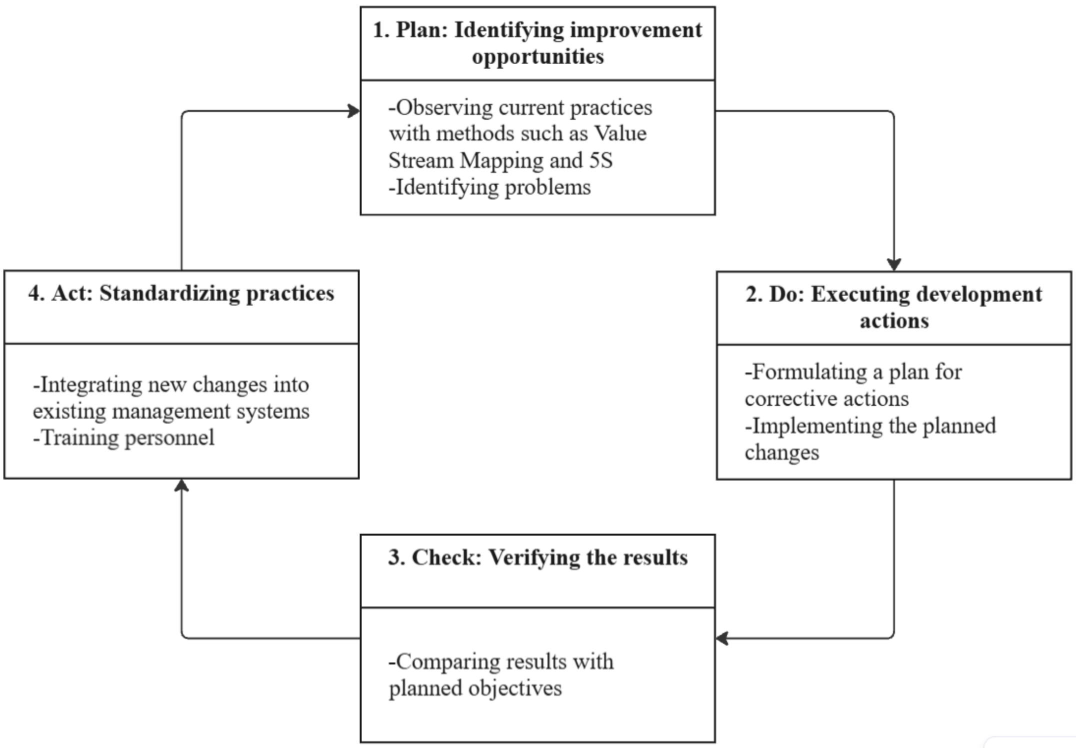

Figure 4: Cycle of continuous improvement activities in a manufacturing context. Adapted from Grünberg (2003).

Despite their widespread adoption, continuous improvement methods are not without critisims. For example, Hopp and Spearman (2011) argue that Lean implementations can result in “blind copying” of best practices without sufficient consideration of context-specific needs. Additionally, while CI methods offer a framework for improving productivity in a manufacturing setting, they do not necessarily comment on "what" a company should improve or "when" an improvement process should happen. To quantify these factors, companies often implement necessary metrics and key performance indicators (KPIs) that help them effectively monitor current processes. The following subchapter discusses these metrics and their relationship to company performance in more detail.

2.3.2 Performance Metrics -based Methods

The previously introduced CI methods represent common managerial frameworks used to guide development actions. However, the effectiveness of these methods depends on accurate visibility into actual operations. This visibility is achieved through the systematic use of performance metrics, which aggregate process-level data into a more interpretable form. As illustrated in figure 3, individual metrics quantify performance factors such as productivity, quality, flexibility, time and cost.

Many of the most commonly used factory-level metrics are ultimately grounded in Little’s Law (Muthiah and Huang, 2006). In essence, they quantify fundamental manufacturing elements such as throughput, work-in-progress inventory, and cycle time through units such as *item count*, *time taken*, or *financial cost*. These unit metrics can then be further combined to create relevant indicators based on the organization’s needs.

Over time, certain manufacturing metrics have become more prevalent than others due to their perceived universality and ease of application. A prominent example of this phenomenon is the ISO 22400 standard (ISO 22400-2:2014, 2014), which specifies 34 KPIs for manufacturing operations management. These KPIs range from fundamental measures, such as the throughput rate, to more specialized indicators focusing on a particular performance factor, such as the quality ratio. The intention behind the ISO standard is to offer a universal reference set for performance measurement that can be adapted and further tailored according to an organization’s needs.

The primary role of these metrics is to support managerial decision making by delivering comparable and actionable insights. This naturally raises the question of how different metrics can be meaningfully compared. From an economic point of view, all performance metrics can ultimately be expressed in monetary terms, as money reflects both the value of manufacturing outputs and the costs of materials and labor consumed in production. Both Misterek et al. (1992) and Schreyer and Pilat (2001) emphasize this cost-based interpretation of performance, arguing that it enables consistent evaluation of resource efficiency and financial return.

However, as discussed in chapter 2.2.1, this economic view is purely based on inputs and outputs, and is often applied at a high level. Due to these reasons, financial metrics themselves rarely provide actionable insight for improving performance on a factory level. In response to these limitations, more granular cost-based methodologies have been developed to bring economic analysis closer to manufacturing processes. Among the most prominent are *Activity-Based Costing* (ABC) and *Value Stream Costing* (VSC). ABC allocates overhead costs based on the actual consumption of resources by individual process activities, resulting in a more accurate picture of the cost drivers in a manufacturing system (Almeida and Cunha, 2017). VSC, on the other hand, aligns more closely with the principles of Lean Production by associating costs directly with value-adding and non-value-adding activities across a product's value stream (Ruiz-de-Arbulo-Lopez et al., 2013). These methods are useful in quantifying the costs associated with different parts of the manufacturing process, providing a clearer basis for improvement activities by enabling comparisons between process steps or value streams. However, both ABC and VSC remain narrowly focused on financial cost and do not account for other performance factors such as time or flexibility.

One widely used metric that aims to bridge the gap between individual metrics and performance factors is *Overall Equipment Effectiveness* (OEE), introduced by Nakajima (1988). OEE is an indicator that aims to quantify the utilization and effectiveness of manufacturing equipment. It is expressed as a value between 0 and 1 and calculated using the following formula:

$$\text{OEE} = A_{eff} \times P_{eff} \times Q_{eff} \tag{2}$$

Where A_{eff} represents availability efficiency, capturing downtime caused e.g. by machine breakdowns; P_{eff} represents performance efficiency, capturing productivity losses caused e.g. by idle time or reduced processing speed; and Q_{eff} represents quality efficiency, capturing losses caused e.g. by product defects. These three components of OEE can be further calculated through operational metrics such as processing rate and defect rate.

As a performance metric, OEE has many strengths compared to simple unit metrics such as *cycle time* or high-level productivity measures such as *total productivity*. It combines the performance dimensions of *time*, *productivity* and *quality* into a single metric, making it particularly useful for operational diagnostics and benchmarking. Due to this, OEE has been adopted as the key metric of *Total Productive Maintenance* (TPM), another common CI methodology used in manufacturing (Muthiah and Huang, 2006).

Despite its strengths, OEE is limited by the fact that it was originally developed to measure individual manufacturing machines. Therefore, depending on the context, it can be too granular of a metric that does not account for interplay between multiple machines or process steps. As a result, its direct applicability can be limited in manufacturing organizations with complex processes. To address this limitation, similar but broader indicators were proposed. Scott and Pisa (1998) introduced the *Overall Factory Effectiveness* (OFE) metric, which was designed to quantify the performance of entire manufacturing plants operating in the semiconductor industry. This metric combined OEE with multiple other quantitative metrics, essentially creating a prototype of the more modern KPI dashboards used commonly in manufacturing management. However, the OFE-metric was not popularized beyond semiconductor manufacturing, and authors such as Muthiah and Huang (2006) criticized it for not being useful for diagnosing problems within the factory, such as identifying bottlenecks.

To further address the challenge of factory-level measurement, Muthiah and Huang (2007) introduced the metric *Overall Throughput Effectiveness* (OTE), designed to extend the logic of OEE from the machine level to the plant level. OTE compares the actual output of a factory to its theoretical maximum output over a given time period:

$$\text{OTE} = \frac{\text{Actual throughput (units) from factory in total time}}{\text{Theoretical throughput (units) from factory in total time}} \tag{3}$$

The exact mathematical formulation of OTE was also developed, but it only extended to four simplified manufacturing subsystems: series, parallel, assembly and expansion. Because of this, Muthiah and Huang (2007) acknowledged that OTE lacked the ability to model an entire factory due to not capturing equipment connectivity information.

Metrics such as OEE, OFE, and OTE exemplify the fundamental challenge in manufacturing performance measurement: the trade-off between scope and specificity. Metrics designed for local precision, such as OEE, may lack broader system-wide insight, while more comprehensive indicators often sacrifice diagnostic clarity or require impractical levels of data and modeling capabilities.

This phenomenon is further complicated by the fact that most of these metrics focus primarily on machines and physical assets, while largely overlooking the human dimension of manufacturing performance. In practice, human elements often play a critical role in the performance of manufacturing systems, as many industrial environments still rely heavily on human labor instead of fully autonomous production systems. Metrics such as OEE are not suited to account for factors such as the *efficiency of labor utilization*, *workload distribution* or *personnel safety*, creating potential blind spots in performance evaluation.

Due to the limitations of individual performance metrics, modern manufacturing organizations increasingly rely on collections of KPIs designed to monitor specific aspects of the production system, which are often visualized through dashboards and integrated into MES or ERP platforms (Xu et al., 2024). Despite these practical advances, companies frequently struggle to identify metrics that are the most relevant to their specific measurement needs. Although these needs are ultimately case-specific, they depend on both the strategic needs and the operational realities of the organization.

2.4 Strategic and Operational views on Performance Measurement

The reviewed literature on the operative nature of discrete manufacturing and performance measuring approaches provides an overview of three key areas related to performance measurement:

1. The concept of performance and the factors that comprise it.
2. A variety of methodologies for performance improvement.
3. Metrics developed for quantifying specific performance factors.

Despite the extensive research conducted in each of these domains, there is no established methodology that clearly instructs organizations on what aspects of their operations should be measured, how those metrics should be selected, or why certain indicators are more valuable than others in a specific context. This lack of guidance increases the risk that performance improvement efforts become generic imitations of industry best practices without a clear connection to the organization's own strategic priorities, as cautioned by Hopp and Spearman (2011).

Adding further complexity, the operational characteristics of manufacturing organizations vary widely across industries, production models, and organizational structures. Consequently, no single set of performance metrics or improvement tools can be universally applied. In order to further understand how performance metrics should be constructed in practice, the following subchapters examine both strategic and operational factors that have the strongest impact in shaping the performance metrics used by an organization.

2.4.1 Strategic Factors Affecting Performance Measurement

The relationship between performance metrics and organizational strategy has been extensively analyzed in academic literature. Authors such as Neely et al. (1995) and Melnyk and Stewart (2004) emphasize the strong interdependence between the two concepts, stating how performance measurement plays a key role in translating an organization's strategy into reality. This is further elaborated on by Henri (2006), who demonstrates how the use of performance metrics helps manufacturing organizations to build their strategic capabilities in four major areas: innovativeness, organizational learning, market orientation and entrepreneurship. While performance metrics alone do not directly determine organizational performance, they are widely regarded as an essential tool providing the foundation for strategic development.

This relationship between strategy and performance measurement has also been described as a two-way interrelationship. In their empirical study on the topic, (Kober et al., 2007) display that performance measurement both shapes and is shaped by strategy. They argue that an organization's strategic orientation provides the need for certain performance metrics, and these metrics in turn provide increased visibility and knowledge, which helps the organization in aligning its strategic goals.

A milder form of this two-way interrelationship can also be encountered in financial indicators, such as those used in the financial reporting of an organization. As strategic choices influence a company's success, the change is reflected in financial metrics. These changes are then used to evaluate the success of the strategic decisions. However, authors such as Neely et al. (1995) argue that financial metrics should not act as primary drivers of strategy, as they tend to be lagging and too aggregated for guiding decisions. Therefore, performance metrics are deemed to have a stronger interrelationship with organizational strategy compared to financial reporting metrics. These relationships between the three concepts are presented in figure 5.

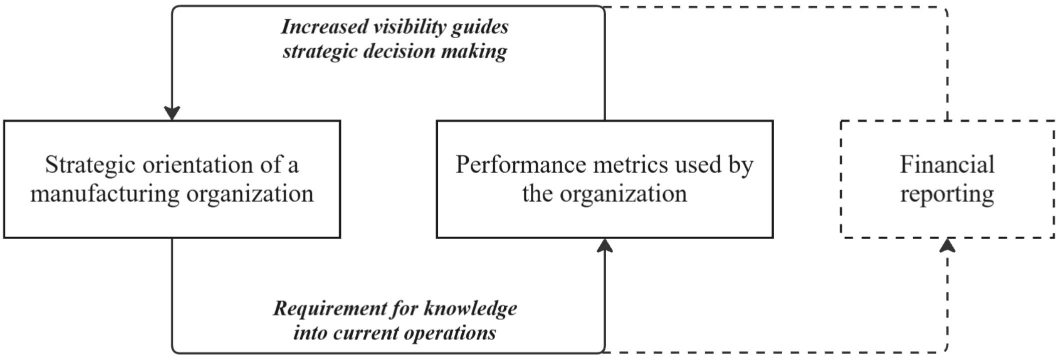

Figure 5: The two-way iterative relationships between strategy, performance measurement and financial reporting. Adapted from Kober et al. (2007) and Neely et al. (1995).

This view drives a separation between financial reporting and other performance metrics of an organization, further challenging the economic view of productivity presented by Misterek et al. (1992) and Schreyer and Pilat (2001). Instead, authors such as Neely et al. (1995) consider economic metrics more as a result of conducted strategic decisions, and emphasize the larger role of company-specific metrics in organizational decision making.

Another factor that further affects the relationship between company strategy and performance metrics is the role of the metrics used. Performance metrics are commonly categorized as either *lagging* or *leading* (Walaski, 2020), depending on the causal relationship they describe. Lagging metrics are indicators that reflect an earlier change in the measured phenomenon, and leading metrics are indicators that are used to predict future changes, giving information on future performance outcomes (Manuele, 2009). In manufacturing, examples of lagging metrics include e.g. *defect rate*, *OEE*, and the aforementioned financial reporting metrics. Examples of leading metrics include e.g. *on-time material deliveries*, which is used to predict material availability and ultimately evaluate the production rate of a manufacturing plant in the near future.

This separation is highly relevant for strategy, as leading metrics can be especially valuable for advancing strategic goals by enabling proactive action. Wilkes (2005) presents multiple examples of organizations deliberately aligning their measurement systems with leading practices to improve their strategic decision making processes. In contrast, lagging metrics often provide a more concrete, standardized basis for tracking the execution of an existing strategy. This is reflected in the fact that the ISO 22400 standard (ISO 22400-2:2014, 2014) consists entirely of lagging metrics aimed at providing operational clarity. However, identifying effective leading indicators is considerably more challenging. As noted by Zheng et al. (2019), leading metrics are often highly company-specific and may require data that is not readily available.

Additionally, the iterative nature between strategy and performance measurement is supported by the fact that it can help in alleviating the blind spots in measuring practices. In their article, Melnyk and Stewart (2004) highlight the frequently mentioned proverb

"You get what you inspect, not what you expect." Overemphasis on a narrow set of metrics can distort organizational priorities, leading to short-term optimization at the expense of long-term strategic guidance.

2.4.2 Operational Factors Affecting Performance Measurement

While strategic objectives set the overarching direction for performance measurement, operational factors determine the feasibility, granularity, and practical implementation of the selected metrics. In discrete manufacturing, these factors are rooted in the physical and technological realities of the production system, which strongly influences both what can be measured and how the collected measurement data can be utilized.

At a high level, the operational factors are largely constrained by where the manufacturing organization is positioned within the product-process matrix (presented in figure 1). This is because variables such as product variety, production volume and degree of automation strongly shape the design of measurement systems. For example, highly automated continuous flow lines can enable the real-time tracking of OEE at a machine level, whereas low-volume, high-mix environments may require more flexible and context-specific metrics that capture variability in setup times or changeover efficiency.

As discussed in chapter 2.1, the visibility into the manufacturing process itself can often be another limiting factor for performance measurement. Yuan et al. (2013) emphasize that transparent information flows in discrete manufacturing are enabled by sufficient data availability and quality. While Industry 4.0 technologies can significantly enhance this transparency (Kamble et al., 2020), their benefits for performance measurement are not automatic. The implementation of new technologies into existing manufacturing systems requires significant time and effort. Additionally, the adaptation of new data streams into existing measurement architectures needs to be carefully planned, so that the resulting information can be translated into actionable insights.

Operational factors also extend to workforce-related aspects. While many measurement frameworks focus on aspect such as equipment or throughput efficiency, the utilization of human resources also remains integral to manufacturing outcomes. Measurable aspects such as *labor utilization*, *workload distribution* or *safety performance* can significantly influence manufacturing performance, yet they remain underrepresented in academic research. In their literature review on the topic of performance measurement, Gomes et al. (2004) found that only 3.9% of reviewed articles focused on human resources. Efforts to integrate human-related metrics into manufacturing performance frameworks have been proposed by authors such as Shahrokhi and Bernard (2009), who developed a model to assess both human productivity and safety risks in production environments. However, the extent to which such metrics are adopted varies widely between industries, and there is currently no clear understanding of how comprehensively human factors are incorporated into common performance measurement practices.

Another major operational factor that affects performance measurement is the intended use case of the adopted metrics. In addition to the aforementioned distinction between leading and lagging metrics, authors such as Brudan (2010) emphasize the categorization between *control* or *learning* metrics. Control metrics are used to monitor whether current processes are operating within acceptable limits, while learning metrics evaluate the impact of improvement initiatives and guide iterative development.

This distinction highlights the differences in the operative use of metrics. As Melnyk and Stewart (2004) note, control metrics enable managers and workers to evaluate and control manufacturing performance on a day-to-day level, whereas learning metrics are used to identify gaps between current performance and expectations, providing direction for long-term operational development. The degree to which a company should balance control and learning metrics in their operations is highly dependent on chosen methods for performance improvement. As stated in chapter 2.3, managerial frameworks such as Lean Management heavily emphasize CI and iterative learning, which naturally increases the relevance of learning metrics in the measurement system. In contrast, organizations operating in highly standardized, compliance-driven environments may rely more heavily on control metrics to ensure stability and adherence to established procedures.

2.5 Theoretical Synthesis and Framework for Evaluating Performance Metrics

This chapter synthesizes key factors identified in the reviewed literature into a theoretical framework that illustrates how strategic and operational considerations jointly shape performance measurement systems. As discussed in the previous chapters, a manufacturing company's strategy ultimately addresses *what* should be measured and *why*, while its operational elements determine *how* and to what extent these measurements can be effectively implemented and used in organizational decision making. Both elements consist of multiple factors whose relative importance should be weighted according to the company's specific needs and context.

Based on the reviewed literature, it is evident that performance metrics and performance improvement activities should not be viewed as separate concepts, but rather as interconnected components of the same system. This partly challenges the perspective of Muthiah and Huang (2006), who describe metrics and improvement methodologies as distinct approaches to improving manufacturing performance. Authors such as Cua et al. (2001) offer a different interpretation: they argue that the selection of a performance improvement method should be considered a strategic decision for a manufacturing organization, because the methods are inherently linked to strategy-oriented practices such as leadership and employee involvement. Additionally, they highlight how companies are not limited to a single methodology. On the contrary, manufacturing organizations adopting multiple improvement methods are often able to leverage synergies between them, ultimately leading to improved company performance (Cua et al., 2001).

While performance improvement methods offer guidance on what type of metrics should be emphasized, they do not eliminate the need to consider the practical realities of manufacturing. In practice, operational factors tend to play a more limiting role by defining the technological, informational, and resource constraints within which strategic ambitions must operate (Hopp and Spearman, 2011). These constraints determine not only which performance metrics can be realistically applied on a factory level, but also the reliability with which they can be measured.

Figure 6 presents a conceptual framework synthesizing these insights. It illustrates how both strategic and operational factors jointly shape the set of performance metrics a company employs.

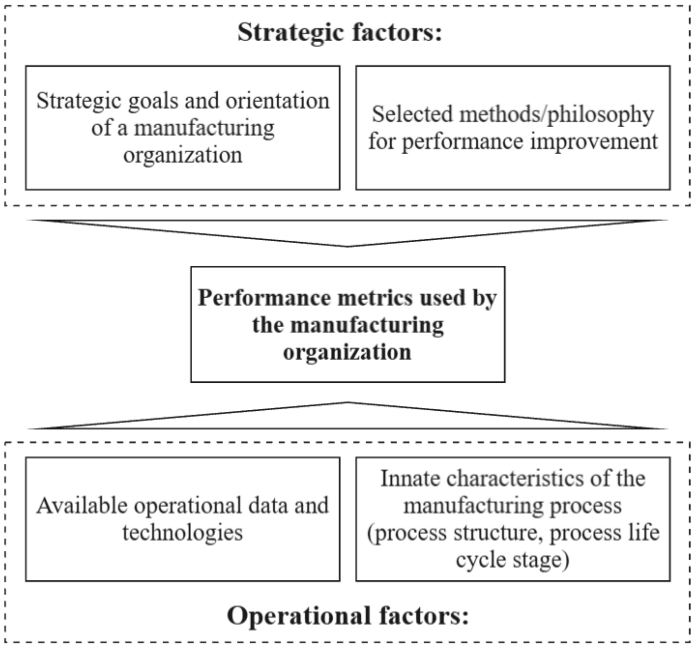

Figure 6: Conceptual framework for aligning performance metrics with key strategic and operational factors in manufacturing organizations.

In the context of this thesis, the role of the presented framework is not to prescribe target performance levels or define the "ideal" measurement system, but to serve as a tool for examining the factors that most strongly influence the formulation of performance metrics in manufacturing organizations. The framework thus focuses on understanding why certain metrics emerge, which factors guide their selection, and how strategic intent and operational realities interact in shaping them. The framework also highlights how the design of performance metrics is always case-specific, reflecting both the strategic orientation of the company and the constraints of its operating environment.

The early proponents of strategic performance measurement such as Kaplan and Norton (1992) emphasized that performance measurement should cover objectives equally across four distinct perspectives (financial, customer, internal and learning). The framework presented here partially challenges this assumption by underscoring the importance of weighting performance factors in accordance with strategic priorities and operational constraints, rather than applying an equal balance by default. This perspective reflects the context-specific nature of performance measurement and the need to tailor systems to the realities of each manufacturing environment.

The subsequent case study of this thesis builds upon this theoretical synthesis by extending the framework into an empirical setting. Whereas the framework itself concentrates on the formation of metrics, the case study examines the application of these metrics in practice. The framework deliberately excludes the decision making dimension presented earlier in figure 5, as its focus lies in identifying the key characteristics that influence the formulation and use of performance metrics, rather than the managerial processes through which measurement results are acted upon.

Ultimately, there is no universal “silver bullet” set of metrics that should be used for improving company performance in a manufacturing context. Therefore, understanding the interaction between strategic and operational factors provides a structured foundation for designing measurement systems that are both relevant and actionable.

3 Research Methodology and Design

This chapter outlines the research design of the study and the methodological choices made to address the research questions introduced in chapter 1.3. The first subchapter discusses the overall research strategy, its underlying methodological assumptions and limitations. The second subchapter describes the design of the single-case study approach and the methods used to ensure both ethical and data security practices. The third subchapter details the data collection process and analysis methods used to generate and interpret the empirical findings. Finally, the fourth subchapter presents the progression of the data analysis process and the results produced after examining and structuring the collected material.

3.1 Research Approach

In addition to the review of academic literature, this thesis employs a case study research approach to analyze the operations of the case company, following the methodological guidelines outlined by Eisenhardt and Graebner (2007) and Yin (2014). A single-case study was chosen, as it is particularly well suited for providing an in-depth, holistic investigation of a phenomenon within its real world context. While multiple-case studies typically offer a stronger basis for building new theory, the purpose of this research is to apply existing theories to the operational context of a single manufacturing organization.

In line with the case study approach described by Yin (2014), the objective of the empirical study is to examine performance metrics as a bounded phenomenon within a clearly defined context: the production operations of a discrete manufacturing company. The study aims to generate results that are both practically actionable (in the form of concrete improvement opportunities) and analytically generalizable, by linking case findings to broader theoretical discussions on performance measurement. According to the categorization of Ketokivi and Choi (2014), this goal classifies the study as qualitative, even though quantitative data sources are also incorporated (as further discussed in chapter 3.3).

More specifically, the case study investigates how the theoretical framework developed through the literature review manifests itself in the practices of the case company. The approach is therefore consistent with what Ketokivi and Choi (2014) define as *theory elaboration*, in which an existing theoretical framework is applied to an empirical setting with the aim of refining, extending, or contextualizing it in light of observed evidence. This differs from other case research emphases: *theory generation*, which develops new constructs directly from empirical data, and *theory testing*, which focuses on validating predefined hypotheses (Ketokivi and Choi, 2014).

To further clarify and structure the methodological foundations of the study, the research approach is mapped onto the widely used ‘research onion’ framework introduced by Saunders et al. (2009), which is used to visualize the explicit methodological choices used in academic research:

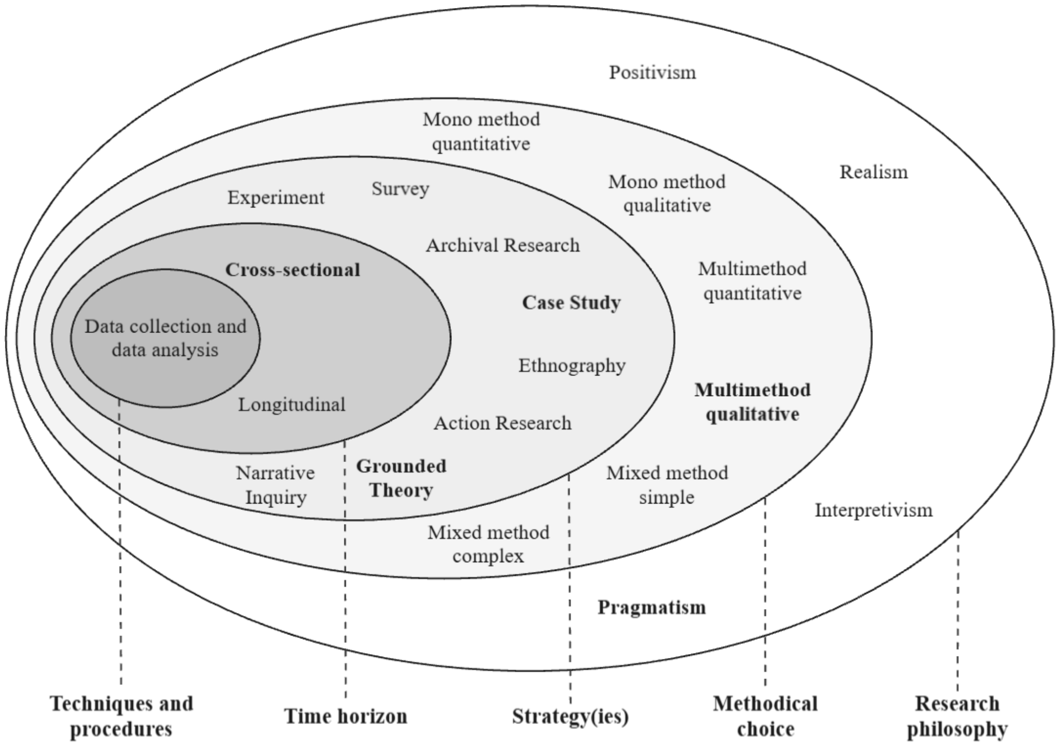

Figure 7: Research onion as presented by Saunders et al. (2009). The methodological principles chosen for the research are bolded.

The research onion is used to illustrate how the data collection and analysis techniques (the core of the onion) need to be considered in relation to other design elements (the outer layers of the onion) (Saunders and Tosey, 2013). The methodological choices made in this thesis can be summarized as follows:

- **Time horizon:** A cross-sectional design was adopted, as data was collected within a relatively short period of time rather than longitudinally.

- **Research strategy:** In addition to the previously outlined case study approach, the study draws on elements of grounded theory as described by Birks and Mills (2012), particularly in the iterative collection of qualitative data.

- **Methodical choice:** The study employs a multimethod qualitative design, combining multiple qualitative data collection techniques (e.g., semi-structured interviews, document analysis, and factory observations) without a heavy emphasis on advanced qualitative data analysis procedures. The details of data collection and analysis are further discussed in chapter 3.3.

- **Research philosophy:** The research is guided by pragmatism, where importance is placed in the practical consequences of the findings. Although case studies are often associated with philosophies of interpretivism and positivism (Saunders and Tosey, 2013), this study emphasizes practical consequences over strict adherence to either paradigm. Pragmatism acknowledges that no single viewpoint provides a complete understanding of a problem, and instead values approaches that generate actionable insights for both practitioners and scholars.

As with all case studies, the chosen methodology also involves certain limitations. The focus on a single case company inherently restricts the statistical generalizability of the findings, as the results are shaped by the technological, organizational, and cultural characteristics of the company (Eisenhardt and Graebner, 2007). However, the objective of this thesis is not statistical generalization but rather *analytical generalization* (Yin, 2014), where the insights from the case can be transferred to other discrete manufacturing contexts facing similar challenges. Another limitation arises from the bottom-up nature of the research approach (further discussed in section 3.2). While the approach enables a detailed understanding of operational realities, it may place less emphasis on higher-level strategic considerations. Nevertheless, this trade-off is essential for uncovering the practical constraints that ultimately shape the effectiveness of performance measurement systems.

3.2 Design of the Case Study

The case company examined in this thesis was selected based on the needs identified during earlier phases of the TwinFlow research project. As the project itself focuses on strengthening the competitiveness of the Finnish industrial sector, a central requirement was to improve the understanding of organizational performance measurement practices and their associated improvement opportunities. From this perspective, the chosen case company provided a particularly suitable research setting, as it had a clear need to refine its factory-level performance metrics and related processes. This alignment ensured that the study would contribute both to the project objectives and to the research questions of this thesis.

The unit of analysis in this study consists of the singular manufacturing plant of the case company. More specifically, the study examines *how* performance metrics are defined and applied in this manufacturing plant, and *why* they are currently used in connection with factory-level processes. The scope of the case study was limited to the company's manufacturing and intralogistics operations, as these areas provide the most direct link between operational data, performance outcomes, and higher-level strategic objectives.

The empirical progression of the case study followed a bottom-up design. In practice, this meant beginning with a detailed investigation of the operative context of the case company, with particular attention to how performance is measured and interpreted on the factory floor -level. This operational perspective was then connected to the company's overall performance objectives in order to identify potential misalignments between strategic intent and practical implementation. By starting from the realities of daily operations, the approach helped to highlight perspective differences between various organizational actors of the case company, including managers, engineers and operators.

In addition to the primary investigation of the case company, the study was enriched by triangulation interviews with representatives of other OEM companies participating in the TwinFlow project. These interviews complemented the literature review by providing external perspectives and examples of best practices from organizations with comparable operations. As emphasized by Eisenhardt and Graebner (2007), such supplementary interviews help to mitigate bias and increase the depth of the analysis. This approach is also consistent with Yin (2014), who highlights that single-case studies should rely on multiple sources of evidence to strengthen the validity of their findings.

The progression of the case study is illustrated in figure 8. The process began with a three-week data collection phase at the premises of the case company (step 1), followed by a review of the collected material and an evaluation of development needs (step 2). In parallel, triangulation interviews were conducted with representatives of other OEM companies to contextualize the findings (step 2.1). Based on the identified needs, potential development directions were formulated and iterated together with the case company in dedicated workshops (step 3). These discussions were further supported by the utilization and elaboration of the theoretical framework derived from the literature review (step 3.1). Finally, the improvement proposals were refined and validated in workshops with company representatives (step 4).

The progression of the case study was therefore designed to both generate and refine improvement ideas through iterative interaction with the case company. This combination of triangulation, utilization of the theoretical framework, and workshops ensured that the study followed the design principles of case study research as outlined by Yin (2014), who describes case research as a "linear but iterative process." In this study, the results of data collection were continuously shared, reflected upon, and revised in collaboration with project participants, thereby enabling iteration while enhancing the robustness of the findings.

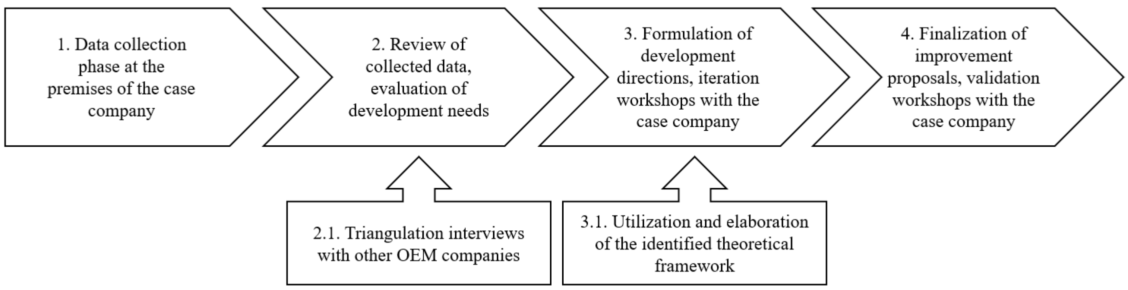

Figure 8: Design and progression of the case study.

Because the empirical research involved human participants, ethical considerations were also carefully taken into account throughout the research. Preventive measures such as informed and voluntary consent were implemented in accordance with the recommendations of Roller and Lavrakas (2015). All interview and workshop data was treated confidentially, and the reporting of results avoids disclosing sensitive company-specific details. Furthermore, numerical data collected at the case company is not presented in its original form in this thesis. Instead, the values presented in chapter 4 are modified or aggregated to preserve confidentiality while still enabling meaningful analysis.

The following chapter outlines the data collection and analysis procedures in greater detail.

3.3 Data Collection

The empirical research relied on a combination of qualitative and quantitative data sources to gain a comprehensive understanding of the case company's performance measurement practices. The primary data collection took place during a three-week period at the company's premises and included semi-structured expert interviews (subchapter 3.3.1), complemented by hands-on observations, internal company data, and relevant documentation (subchapter 3.3.2). In addition, two external semi-structured interviews were later conducted with representatives of other OEM companies to provide comparative perspectives (subchapter 3.3.1).

The interview data served as the main source for addressing the research questions and for structuring development suggestions for the case company. The complementary data sources functioned as secondary material, supporting the analysis and strengthening the validity of the findings.

3.3.1 Semi-structured Expert Interviews

The research combined semi-structured interviews with professionals employed both by the case company and by other selected OEM companies. Within the case company, interviewees were chosen according to two main criteria: (1) their knowledge of current production and IL processes, and (2) coverage across different organizational levels and domains. These criteria were used to ensure multiple perspectives on the research topics and to minimize potential blind spots in data collection. The initial interview participants were drawn from staff members who had already expressed willingness to take part in the project. Subsequently, snowball sampling was applied, whereby recommendations from earlier interviewees were used to identify further participants.

These interviews were mainly conducted during the on-site data collection period. In total, 10 interviews were carried out with 11 employees of the case company, primarily targeting managers and specialists involved in production, quality, logistics, and data management. The interview details are presented in table 1, and the interview structure is presented in appendix A. The interviews were organized jointly with another Master's thesis worker participating in the TwinFlow project, due to which the interviews included questions related to both thesis topics.

Interviews were conducted in Finnish, both in-person and via Microsoft Teams. In-person interviews were recorded using a mobile device, while online interviews were recorded using Teams' built-in recording functionality. All recordings were automatically transcribed using Aalto University's *Speech2Text* software and then manually corrected in collaboration with the other thesis worker. The finalized transcripts were imported into the *ATLAS.ti* software platform, where they were coded and analyzed following the qualitative research method described by Gioia et al. (2013). The results of this analysis are presented in chapter 3.4.

At the beginning of each interview, informed consent was obtained by requesting voluntary participation. Participants were also asked for permission to record and transcribe the session, and their anonymity was assured.

Table 1: Interview details: internal interviews

Interviewee number	Role	Interview length
1	Process Owner, Quality NPD	41:52
2	Systems Specialist, Logistics	1:35:30
3	Senior Manager, OPEX	1:14:20
4	Senior Manager, Quality	35:27
5	Production Manager (1)	42:43
6	Factory Manager	49:18
7	Data, API and AI Manager	52:59*
8	Data Architect	52:59*
9	Production Manager (2)	49:43
10	Systems Specialist, Operations	1:29:54
11	Manufacturing Engineering Manager	1:19:02

* Interviewees 7 and 8 participated to the same interview.

In addition to the internal interviews, two semi-structured interviews were later conducted with representatives of other OEM companies participating in the TwinFlow project. The interviewees were identified through the project network: first by approaching TwinFlow consortium members for suitable professionals, after which the companies themselves nominated domain experts with relevant experience. One of these interviews was conducted remotely via Microsoft Teams, while the other took place on-site and was conducted in collaboration with another Aalto University's research team. A slightly adapted interview guide was used for these discussions, which is provided in Appendix B. This interview guide focused more closely on the topic of productivity measurement, as the limited time spent on external interviews did not allow for in-depth analysis of the entire performance measurement system.

These external interviews offered complementary perspectives and examples of best practices from organizations operating in comparable manufacturing environments. Both companies operate assembly manufacturing plants in Finland, making them particularly relevant points of comparison for the case study. The details of these interviews are presented in table 2. The interviews were recorded, transcribed, and analyzed using the Gioia method, but their results were not aggregated into the same coding structures as the internal interviews. Instead, their insights are referred to in chapter 4.2, where they are used to contextualize and validate the proposed improvement opportunities.

Table 2: Interview details: external interviews

Interviewee number	Role	Company	Interview length
E1	Senior Production Development Manager	Mitsubishi Logisnext	1:01:22
E2	Plant Director	Konecranes	2:03:33
E3	Plant Manager	Konecranes	2:03:33

* Interviewees E2 and E3 participated to the same interview.

3.3.2 Complementary Information Sources

In addition to the internal interviews, complementary data was collected during the 11-day on-site data collection period. This included personal notes based on hands-on observations of daily operations collected e.g. through informal discussions with factory personnel, which provided contextual understanding of the production environment and actual processes.

Further information was obtained from the internal IT systems of the case company, including Manufacturing Execution System data, Enterprise Resource Planning records, and Power BI reporting. Due to confidentiality, the raw data from these systems is not displayed in this thesis. Instead, selected aggregated/modified results are referenced where relevant to the analysis. To ensure data security, all information was primarily accessed through laptops owned by the case company, and only limited extracts were shared for further analysis.

Finally, internal documentation such as safety materials and assembly instructions was reviewed. These documents offered additional insight into the company's operational practices and supported the interpretation of interview and observational data. An overview of these complementary data sources is provided in table 3.

3.4 Data Analysis

Once the data had been collected, it was analyzed through a structured qualitative process. The data analysis process utilized case study tactics highlighted by Yin (2014) to ensure internal validity. This included applying cognitive mapping and explanation building to the collected data as a way to establish causal relationships and connections between the studied phenomena. Consistent with this approach, the analysis was designed to balance theoretical guidance with openness to unexpected findings. The literature review served as an interpretive foundation for the analysis, while the collected material was used to provide contextual depth and identify new themes. This approach was used to serve the theory elaborating approach of the research.

The core of the empirical analysis was based on the transcripts of the internal expert interviews. After the internal interviews were transcribed in Finnish, they were analyzed using the qualitative research methodology introduced by Gioia et al. (2013), which was selected due to its structured approach for developing inductive

Table 3: Descriptions of complementary information sources

Data source	Description
Hands-on observations	Notes collected during the on-site data collection phase, based on informal conversations with factory personnel, shadowing of manufacturing and assembly tasks, and limited participation in selected assembly activities and internal meetings.
Operational and reporting data	Data extracted from the company's internal systems, including productivity metrics, production disruption records, safety event logs, and machine-level measurements, inspected during the on-site data collection phase.
Internal documentation	Company documentation in both digital and paper formats, including general safety and assembly instructions as well as product- and part-specific materials related to inventory management and factory flows.

findings from qualitative data. In practice, this first included identifying relevant emerging opinions from the transcribed interviews, and coding them as *first-order concepts*. Initially, approximately 95 first-order concepts were identified from the internal interviews. Subsequently, these concepts were reviewed and aggregated based on their content, since many of them contained overlapping information and highlighted similar phenomena. After reviewing the first-order concepts and their underlying interview quotes, 37 final first-order concepts emerged from the interview data.

These concepts were then further aggregated into 12 different *second-order themes*, focusing on the theoretical implications of the identified concepts (Gioia et al., 2013). In the context of this research, the second-order themes were aligned with themes of performance measurement identified during the literature review. Finally, the 12 themes were organized into three aggregated dimensions, each corresponding to one of the research sub-questions introduced in chapter 1.3.. The resulting data structures are presented in tables 4, 5, and 6. This systematic process was used both to ensure transparency and to highlight the thought process behind the qualitative analysis. The findings from this analysis are further discussed in chapter 4.

In addition to the interview transcripts, complementary materials were analyzed in a more informal but systematic manner.. Notes from hands-on observations, internal documentation, and selected system data were used to both form a clear picture of the case company's current operations and to validate and enrich the findings emerging from the internal interviews.

Table 4: Gioia data structure for SQ1, displaying strengths and limitations of the existing metrics used by the case company

First-Order Concepts	Second-Order Themes	Aggregated Dimension
High-level metrics include factory productivity, scheduling hold and production disruption count. *PPM metrics used for supplier parts.* *Monitoring production buffers and workforce count assist in daily management.*	Current metrics produce information on multiple performance dimensions	Strengths and limitations of existing metrics (SQ1)
Process completion timestamps are not filled in real time. *NPD projects not accounted for in current metrics.* *PPM not suitable for all suppliers.*	Discrepancies between operational realities and information produced by metrics	
Theoretical workload is unknown for some tasks. *Human contribution not currently trackable.* *Product-level visibility is limited.*	Operational factors left out by metrics	
Goal levels in current metrics remain vague. *Current productivity metric can not be used in knowledge-based management.* *Takt time is the primary driver behind productivity metric.*	Unclear use cases for current metrics	

Table 5: Gioia data structure for SQ2, displaying key improvement areas in performance measurement practices of the case company.

First-Order Concepts	Second-Order Themes	Aggregated Dimension
NPD projects are not taken into account in current metrics. *Machine-level data would allow analysis of both costs and improvement opportunities.* *Current metrics only differentiate between machines and harvester heads.*	Distinguishing between different projects & end products	Key improvement areas in current performance measurement practices (SQ2)
Goal is to divide work load as evenly as possible. *Cost dimension is not currently considered in operative metrics.* *Real time workload data would be valuable for production planning* *"Value" of most production steps remains vague.*	Need for better decision support tools	
Variations in daily production are not measured. *Workload of individual machines is not currently visible* *Current metrics aggregate data from the whole factory.* *Need for in-person checks to gain manufacturing visibility.*	Transparency into production process steps	
"Rejections" of individual products should be measured in the future. *Live work load data could be used to manage production more efficiently.*	Need for increased granularity for current performance metrics	

Table 6: Gioia data structure for SQ3, displaying data and system capabilities required for improving performance measurement for the case company

First-Order Concepts	Second-Order Themes	Aggregated Dimension
Machine order cannot be changed in MES. *WMS can not be currently used in parts manufacturing.* *External software options sought for capacity planning.*	Feature limitations of current IT systems	Data and system requirements for improving performance measurement (SQ3)
Machines such as welding robots produce data which is not analyzed further. *Human contribution is only monitored through occasional time studies.* *Real time visibility into production load could assist in workforce management.*	Utilization of new data sources	
Causal relationships can not currently be monitored for production disruptions. *Data sources remain fragmented, hindering availability of data.* *IT department is working on a data platform, that aims to provide a "single source of truth" for company data.*	Increased system connectivity	
Data on production disruptions is often incomplete. *Workers do not log in to MES.* *Visibility inside takt time is currently limited.*	Missing or unsatisfactory data	

4 Findings

This chapter presents the empirical findings of the study. The findings are organized into three main parts. First, the current state of the case company's production system and performance measurement practices are described through the frameworks introduced in section 2. Second, the strengths and weaknesses of existing performance metrics are discussed. Third, opportunities for improvement are presented, highlighting both the proposed development directions and the operational requirements for their implementation.

Throughout this chapter, the findings are presented in a descriptive manner in the context of the case company, supported by evidence from interviews and complementary data sources. The discussion of the theoretical implications, significance and reliability of the results is reserved for chapter 5.

4.1 Analysis of the Current Situation

The case company operates a single manufacturing plant located in Vieremä, Finland. The facility, which was significantly modernized in 2018, serves as the sole production site for the company's full range of forestry machinery. Currently, the company produces three different machinery product groups, each with multiple product models. These product groups are described in table 7.

Table 7: Description of the case company's product groups, adapted from Ponsse (ndb).

Product group	Description	Count of models
Harvesters	Purpose-built machines used for felling trees, delimbing, and cutting them into predetermined lengths at the logging site.	7
Forwarders	Load-carrying machines designed to transport felled and processed logs from the harvesting site to a roadside landing area.	10
Harvester heads	Modular cutting units mounted to harvesters for tree felling. These can also be attached to competitor's harvester products.	12

In addition to the different product models in each product group, each manufactured unit includes a significant degree of customization according to customer requirements. This customization introduces additional complexity into the production process, as different product configurations require specific parts, process adjustments, or additional work stages. The company also produces complementary products, such as information systems for forest machinery and training simulators, but since these are manufactured outside of the Vieremä facility, they are excluded from the scope of this research.

The Vieremä plant employs several hundred people and encompasses both assembly and parts manufacturing processes supported by intralogistics. As a recently modernized facility, it makes use of advanced technologies and generates large volumes of operational data. The company already electronically monitors its factory operations through several performance metrics, which will be further discussed in the following subchapters.

However, from the perspective of the case company, the current metrics in use at the Vieremä plant do not fully support the organization's operational and strategic goals. In particular, the productivity metrics do not sufficiently capture variations in product value, the degree of customer-specific customization, or the degree of subcontracting in production. As a result, the measurement system provides only limited visibility into the true drivers of performance, and opportunities for improvement are not systematically recognized.

Accordingly, the objective of this part of the research is to establish a baseline understanding of the current state of production and performance measurement at the case company. This provides the foundation for developing improved performance metrics and identifying the types of data required to support both immediate operational improvements and longer-term development of factory performance.

4.1.1 Structure of the Case Company's Production System

The production system of the case company follows a made-to-order (MTO) model, in which each manufactured product is specified in advance by the customer. This means that the factory does not aim to maximize output in terms of volume, but rather to optimize the use of resources in fulfilling the customer order backlog. As such, the emphasis of the production system lies in achieving efficiency and flexibility, rather than in producing standardized products at maximum speed.

The internal pace of the production system is guided by takt time, which represents the predetermined time at which products move between workstations. The takt time is strongly influenced by the order backlog of the company, and it serves as the central mechanism for balancing workloads across the different assembly stations by setting the concrete pace at which work-in-progress units move between stations. Factory production is organized into two eight-hour shifts across five weekdays, with overtime shifts occasionally scheduled to balance demand peaks or catch up with production delays.

The Vieremä facility combines both parts manufacturing and final assembly processes, supported by intralogistics operations. The factory maintains a relatively high degree of vertical integration, as critical components such as machine frames are welded on-site from raw steel before being passed on to assembly. The assembly process itself is designed as a connected line flow consisting of individual workstations, each operating under a pull control system. All harvesters and forwarders progress through the same sequence of assembly stations, while harvester heads are assembled in a dedicated line due to their distinct process requirements. The production process is thus characterized by **multiple product types**, **low overall volumes**, and a **high degree of customization** in each manufactured unit.

Figure 9 illustrates the general layout of the manufacturing plant. The facility is divided into two main areas: parts manufacturing (shown in blue) and assembly (shown in green). The assembly area also houses the internal logistics center, which coordinates material flows throughout the plant. Arrows 1–3 in the figure indicate inbound material flows, while arrow 4 illustrates the outbound flow of completed machines, which are sent to testing before delivery. A simplified representation of the overall production flow is marked with a gray arrow.

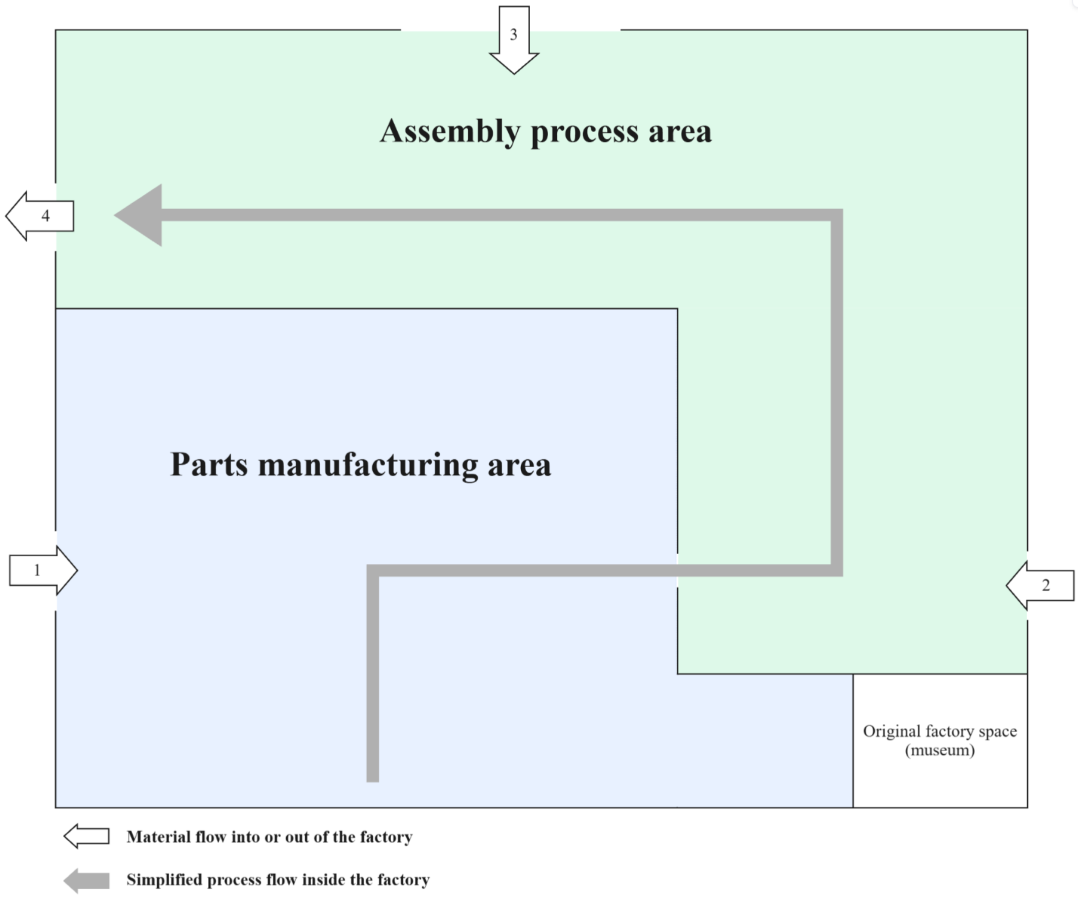

Figure 9: Illustrative layout of the case company's manufacturing plant.

From the perspective of the product–process matrix (see figure 1), the case company's production system is situated between heavy equipment manufacturing and automotive assembly. On the one hand, the factory produces multiple product types in relatively low volumes, which would typically align with a job-shop or batch production environment. On the other hand, the use of a takt time –driven line flow resembles the highly structured processes of high-volume assembly industries. The hybrid positioning is largely explained by the high degree of customization in each manufactured machine, which introduces significant variability into the production process. This variability and the complicated structure of the manufactured products results in the manufacturing process differing from those typically found in the automotive industry. In the automotive industry, lower product variation and a high degree of standardization result in modern assembly lines often operating at takt times of 30–120 seconds (Roser et al., 2025).

This unique production structure creates specific challenges for performance measurement, as metrics must accommodate both the takt time -driven nature of the line flow and the variability inherent in customized heavy equipment manufacturing. The following section examines how performance is currently measured within this production system.

4.1.2 Use of Performance Metrics

To further understand the usage of performance metrics inside the case company, the findings from the empirical study were compared against multiple performance frameworks identified in section 2. At the time of writing, the case company's Vieremä manufacturing plant conducts performance measurement across all of the five performance dimensions identified in figure 3. The measurement system therefore covers productivity, quality, flexibility, time, and cost dimensions of the company's production and IL operations. Based on the interviews and on-site observations, certain metrics emerged as particularly central to the operations of the case company. These are here referred to as the *major performance metrics*. The major metrics are distinguished by their wide and frequent use within the organization to quantify operational factors and support decision making.

An overview of these metrics is presented in figure 10, where they are allocated to the performance dimensions identified in the literature review. The detailed definitions and calculation logic of each metric are further described in appendix C.

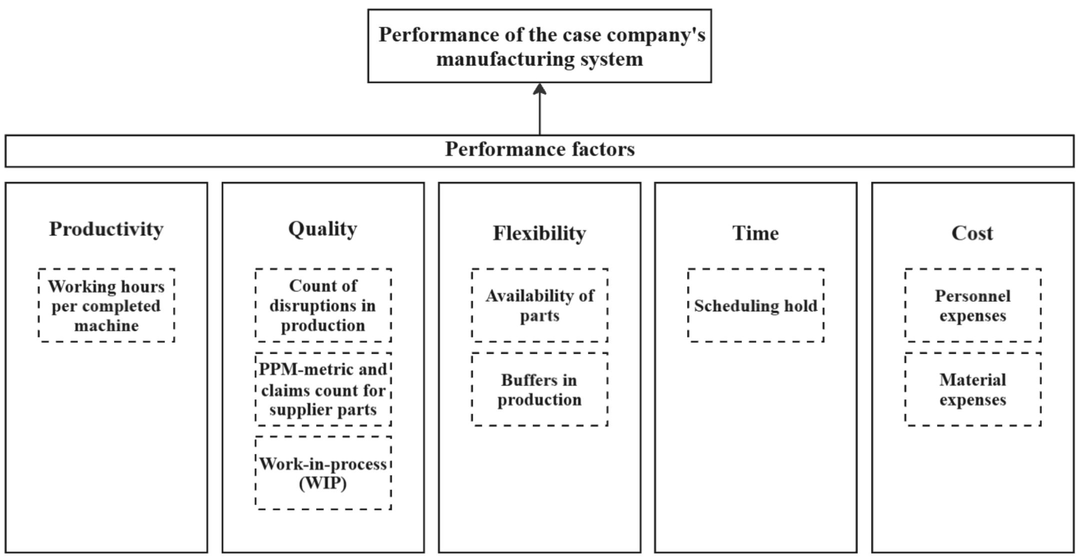

Figure 10: Major performance metrics used by the case company, allocated to performance dimensions identified through the literature review.

While the major performance metrics provide information about several different aspects of the production system, their relative importance differs considerably within the case company. Among these metrics, particular attention in this study was initially placed on the factory-level productivity metric. This measure was consistently highlighted in the interviews as one of the central performance indicators and the one most prone to practical challenges. It is defined as the ratio between the total hours worked in production and the number of finished machines produced by the factory, resulting in an aggregate measure of resource efficiency:

$$\text{Factory productivity} = \frac{\text{Hours worked in production}}{\text{Actual throughput (units) from factory}} \tag{4}$$

In practice, this productivity metric is used by the case company to monitor the efficiency of the manufacturing plant over longer time horizons, typically months or years. Its appeal lies in the fact that it produces a simple, easy-to-understand value that can be used to monitor the competitiveness of the production system. Conceptually, it resembles the Mills Productivity Index, which was introduced in the early 1900's to measure productivity at industry level (Goshu et al., 2017):

$$\text{Mills Productivity Index} = \frac{\text{Output}}{\text{(Number of wage earners)}} \tag{5}$$

In addition to factory productivity, interviewees emphasized the importance of the scheduling hold metric, which measures the deviation between the planned and actual completion time of an individual machine. This is another widely applied measure in manufacturing, defined as follows:

$$\begin{aligned}\text{Scheduling hold} = {} & \text{Scheduled processing time for machine } i \\ & - \text{Actual processing time for machine } i\end{aligned} \tag{6}$$

While scheduling hold is measured in minutes, it is used as a proxy for evaluating whether production remains aligned with customer delivery schedules and financial planning. It can be calculated either at the level of the entire production process or for individual workstations, depending on the required level of granularity. Together, factory productivity and scheduling hold form the two central indicators that guide the operative activities of the case company. As one interviewee summarized:

> *"Well, we have the scheduling hold metric, which is propably the most important one for us. It shows whether we are staying on schedule financially or not. And then there's the previously mentioned "hours per machine" -metric, which has its own shortcomings."*
>
> – *Interviewee 3*

Scheduling hold differs from factory productivity in that it provides more real-time data from current operations, enabling managers to monitor short-term schedule adherence alongside long-term efficiency trends. What makes both of these measures especially valuable is their direct connection to the company's strategic objectives. While the MTO production model means that the company faces no acute pressure to maximize outputs, the company views productivity development as a long-term strategic goal, alongside safety and quality. Improved productivity-related metrics could also provide a stronger foundation for workforce and production planning.

In addition to grouping the major performance metrics under the five identified performance factors, their use in practice can also be understood through two dimensions introduced in chapter 2.4: their orientation toward daily operative management versus strategic monitoring, and whether they function as leading or lagging indicators. This classification, illustrated in figure 11, is based on a synthesis of the collected interview and complementary data and was validated with the representatives of the case company.

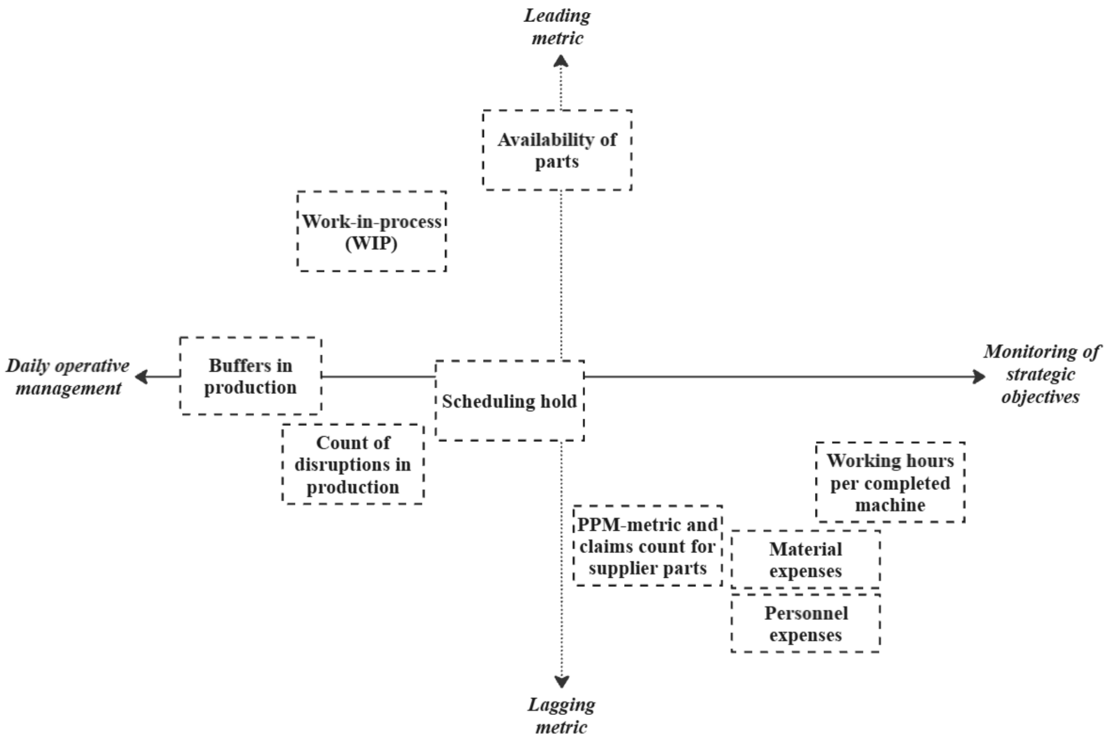

Figure 11: Classification of the case company's major performance metrics by strategic vs. operative usage and leading vs. lagging orientation.

The figure positions each of the major metrics along these two axes to highlight their relative significance within the company's performance measurement system. It illustrates both the practical contexts in which the metrics are applied and whether the metric relies on forward-looking (leading) or backward-looking (lagging) information. For example, the *working hours per completed machine* -metric aggregates monthly production data from operations and is therefore primarily used for high-level strategic monitoring rather than for day-to-day operational control. Conversely, the *count of disruptions in production* functions as an operative tool for managing short-term bottlenecks and relies on near real-time or slightly lagging data (e.g., daily).

The following sections discuss the use cases of the metrics in more detail and analyze the gaps between the case company's measurement needs and the realities of its current performance metrics.

4.1.3 Strengths of Current Metrics

The performance metrics currently utilized by the case company reflect the company's long history in industrial manufacturing and its existing deep understanding of the production process itself. The Vieremä plant already employs several major metrics that are largely consistent with those commonly used in the manufacturing industry (e.g. ISO 22400-2:2014). Key operational factors innate to manufacturing are already visible in the current metrics, as elements such as throughput, WIP levels, and takt time –guided cycle times are all systematically monitored.

This demonstrates that the company has a clear understanding of which aspects of production are worth measuring and how these measures support both operative and strategic objectives. In this respect, the company's practices are aligned with the principles of performance measurement described in the literature, where operational visibility is emphasized as a prerequisite for improvement (Liukkonen and Tsai, 2016).

A second strength lies in the fact that the current metrics are already used to guide the company's continuous improvement activities. The company is an active practitioner of the Lean philosophy and pursues continuous improvement not only through operational practices, but also through the development of its measurement system. A concrete outcome of these development activities is the recent implementation of the MES system throughout the Vieremä plant. The MES enables systematic monitoring of production activities and the collection of large volumes of operational data. In practice, it functions as a knowledge source for production workers by providing access to machine queues and work instructions. Additionally, the MES records operational events, which in turn feeds data for higher-level monitoring and reporting. As one interviewee noted, the current MES supports processes particularly well under stable operating conditions:

> *"Our MES covers the ideal process quite well. But then if problems or other additional process steps arise, they might not be logged into the system and therefore are not visible in reporting."*
>
> *– Interviewee 1*

This observation highlights that the factory is already capable of collecting near real-time data from its production processes and incorporating it into performance metrics, although with certain limitations. In addition to active data collection, another strength lies in how the current metrics are integrated into the company's daily management routines. The case company conducts several recurring daily meetings that revolve around key indicators and tracks continuous development through tools such as Kanban boards. One interviewee described the daily process as follows:

> *"We follow our production daily. We start our mornings by holding an internal quality meeting. After that, we have a team leaders' meeting, where myself, two others, and the quality manager take a look at the big picture. Then there is the morning production meeting, where we usually review schedules, WIPs, and other matters."*
>
> *– Interviewee 1*

The data collection and participation in selected daily meetings revealed that these meetings rely primarily on operative metrics discussed in the previous chapter. This indicates that the distinction between strategic and operative metrics has already been recognized by the case company, reflecting a mature understanding of performance measurement practices.

Taken together, these strengths demonstrate that the case company's measurement system is broad in scope, actively used to support continuous improvement, and well integrated into organizational routines. However, the research also revealed several challenges, particularly in linking the operative and strategic dimensions of performance measurement. These weaknesses are further discussed in the following subchapter.

4.1.4 Weaknesses of Current Metrics

While the current performance metrics provide useful high-level insights, several weaknesses were identified during the study that limit their effectiveness in guiding both operative and strategic decision making. These weaknesses emerged naturally during the data collection process, as interviewees openly highlighted the challenges they encountered in the current performance measurement system. Additionally, practical observations and note taking revealed clear discrepancies between the performance metrics' operational logic and their application in practice. To structure these findings, the identified weaknesses are segmented into four categories.

1. Lack of required level of granularity:
First, several of the existing metrics are highly aggregate in nature, and therefore lack diagnostic capability. Metrics such as *material expenses*, *personnel expenses* and the *factory productivity* currently include data from operations of the entire plant. While such metrics provide aggregate information that is useful for monitoring overall performance over longer time horizons, they cannot be used to pinpoint the underlying causes of positive or negative changes. Several interviewees emphasized this issue, noting that the current figures do not provide sufficient visibility into the production process to support daily management:

> *"We're looking at too big a picture, so maybe we should also pay attention to those individual cells and what they're doing [in the production process]."*
>
> *– Interviewee 1*

> *"Yes, so our productivity figures, whatever productivity figures we have, are at a terribly high level. And I personally don't even want to use them, because they don't tell me the truth."*
>
> *– Interviewee 6*

Other metrics used by the company, such as *buffers in production* or *scheduling hold*, are better suited for granular inspection of daily operations by monitoring events in real time, but they are not currently used to track long-term trends. This suggests a partial misalignment between the needs of the organization and the capabilities of the current measurement system.

This problem is most evident in the case of the factory productivity metric, which provides an overall efficiency measure for the plant but cannot be used to identify bottlenecks or localized issues in the production flow either in the short or the long term. Because the existing metrics do not provide sufficient diagnostic detail, production supervisors are often forced to rely on their own observations and experience when identifying problems on the shop floor.

This indicates that, although the company uses metrics for both daily and long-term management, the current system does not provide adequate information for systematically identifying improvement opportunities in production. Such a gap is misaligned with the company's Lean and continuous improvement philosophy, which emphasizes station-level visibility and iterative development.

2. Blind spots in measurement:
Second, the current metrics contain some blind spots that limit their ability to fully represent performance. Again, interviewees highlighted the factory productivity metric as a clear example. In its current form, all machines are treated equally in the productivity calculation, even though they differ substantially in terms of the work required in production and their final sales value. This is particularly problematic in a made-to-order environment, where product customization is extensive and production content varies significantly from one order to another. Moreover, the numerator of the productivity metric (hours worked in production) is influenced by factors beyond actual workload. For example, the outsourcing of certain production tasks can distort the measure without reflecting real efficiency changes:

> *"[Outsourcing] looks good in the productivity metric: If we were to outsource nearly everything and only put in the last bolt ourselves, then we would be unbelievably efficient in the terms of the "hours per machine" metric. However, that would not likely look good in monetary terms."*
>
> *– Interviewee 3*

This example demonstrates how the productivity metric fails to capture the value-adding activities of the production process. To compensate, the factory organization places greater emphasis on evaluating the *effectiveness* of the work performed. While effectiveness is not systematically measured in the current system, proxy measures are used, particularly through workload monitoring. Workload refers to the degree to which human resources are occupied relative to takt time, and can be expressed as:

$$\text{workload (\%)} = \frac{\text{time spent working during takt time}}{\text{takt time}} \tag{7}$$

Currently, this tracking is conducted through occasional clockings by managerial staff, which provide estimates of workload for each workstation. However, interviewees pointed out that workload is also affected by factors such as the mix of incoming machines and the skill levels of employees present on the line, which decreases the validity of the clocking measurements. Some interviewees suggested that more advanced and systematic workload tracking could unlock better capabilities for workforce planning:

> *"The workload is such a descriptive thing that if we can figure it out, we'll know how many people should actually be working in that department. "*
>
> *– Interviewee 3*

Although a real-time workload measure could provide valuable insights into the flow and efficiency of the production line, it is not currently part of the company's measurement system, as implementing such a measure would require operational capabilities that the company does not yet possess. This issue is revisited in chapter 4.2.

In addition to these gaps, the current performance measurement system also suffers from blind spots that arise from discrepancies between operational realities and the data recorded in the system. In practice, factory personnel sometimes enter information in the wrong format or at the wrong time, which makes the aggregate measures difficult to interpret and occasionally inaccurate. These problems are especially visible in the metrics for production disruptions and scheduling hold. As one interviewee noted:

> *"We monitor [disruptions] digitally, but there are bound to be errors in the information we collect digitally about the manufacturing process, because people are manually entering the data that forms that information."*
>
> *– Interviewee 4*

> *"It bothers me that we are so close to having a great amount of continuous data on this situation. If only all people could be trained, taught, and required to do things the same way [...]. That way, we would accumulate a lot of data points and we would then be able to see product-specific differences and other things."*
>
> *– Interviewee 4*

These inconsistencies therefore arise from current data collection methods and are strongly connected to the human side of operations. Since data entry depends on manual practices and individual interpretation, the quality of information varies and systematic comparability is compromised. As a result, the metrics cannot always be relied upon to provide an accurate representation of the underlying processes.

3. Complexity and deviations are not fully captured:
Third, the existing productivity metrics fail to capture the effects of process complexity and deviations from standard operations. In practice, the plant's production is characterized by variation caused by different product models, frequent disruptions, and the integration of new product development projects. Variability has long been recognized as a factor that negatively affects productivity in manufacturing environments with fixed cycle times (Adams et al., 2001). Yet, in the case company, these sources of variation are not adequately reflected in the current performance measurement system. Instead, they introduce significant fluctuations into the daily workload that remain difficult to evaluate through aggregate metrics. As a result, process improvements or setbacks related to these issues are not systematically captured in the measured performance.

> *"There's bound to be some fluctuation, since our manufactured products differ in their processes. There's no way to make that visible."*
>
> *– Interviewee 1*

A key example of this limitation is the scheduling hold metric, which is currently calculated based on the completion time of the most recently finished product at the final workstation. While this approach provides a simple indicator of schedule adherence, it only captures a single outcome value and therefore reflects the process in a very limited way. To supplement this, the company also monitors scheduling hold at the level of individual workstations. However, this data is currently used primarily for daily problem solving rather than being analyzed statistically to identify long-term patterns.

In reality, deviations occurring earlier in the subassembly or assembly process frequently create bottlenecks that delay subsequent stages. Yet the underlying causes of these delays and their downstream effects remain invisible in the current measurement system. By developing metrics that better account for variability over longer time horizons, the company could build a more accurate understanding of how complexity and deviations influence overall performance. This would also support the principle that variability should first be minimized before processes are optimized further. As one interviewee emphasized:

> *"It is important to understand that things do not necessarily happen according to your plans or how you imagined them to happen. There may be hidden agendas and things that we do not even understand, because our production activities are quite broad. And in that sense, we must constantly strive to find ways to standardize this work."*
>
> *– Interviewee 6*

This highlights that while the company is aware of the need to manage variability, the current metrics do not provide sufficient visibility into its causes or effects. This gap limits the company's ability to systematically address complexity as part of its continuous improvement efforts.

4. Practical limitations in data collection:
Finally, some performance metrics are hindered by limitations in the availability and quality of data. These restrictions reduce the accuracy of the information obtained and diminish the usefulness of metrics for operative decision making.

One important limitation concerns the restricted visibility of product-level data. Although the MES system currently tracks product identifiers, it does not yet provide a comprehensive view of each unit's journey through the factory. As a result, it is not possible to fully trace an individual machine across different stages of the production process. This gap was described by one interviewee as a critical capability that could be addressed in the future:

> *"When product information becomes available to us, it can be used to enhance other data. Perhaps one day, when you ask for the serial number, the entire route through the factory with its various stages and timestamps will be available."*
>
> *– Interviewee 11*

According to factory personnel, this functionality is currently under development at the proof-of-concept stage. Once in place, it could help assess the impact of disruptions on individual products and enable comparisons between different product models.

Another limitation concerns the measurement of human labor. At present, production employees do not log in to the MES system, which makes it difficult to monitor the actual allocation of human resources across different workstations. This creates challenges for evaluating workload distribution and assessing the efficiency of manual work.

> *"If we have a workstation where the assumption is that three employees would be working there, then the fact that we cannot see from the MES whether there are one, two, three, or four employees there, is a problem. This then leads to the difficulty of monitoring human labor, because at the moment, employees do not log in to the MES system in any way, which makes it very difficult to monitor the majority of them."*
>
> *– Interviewee 10*

This limitation is also expected to be addressed in the near future as part of ongoing development projects, as it has potential to assist in evaluating the effectiveness of human labor in the production process and enhance work shift planning. However, in the current state, the absense of this data makes it more difficult to monitor human work.

In addition, smaller but recurring issues were identified in the technical handling of data. Some interviewees noted that MES reports are typically delayed by several hours, which reduces their usefulness for real-time monitoring. At the same time, representatives from the IT department pointed out that development priorities are closely tied to business needs and resource constraints:

"[IT development] really depends on the business need, how and where the data will be used. Well, a good example is for example the disruption data. This data is not worth much if the data arrives a day after the disruption has occurred."

– Interviewee 7

This statement highlights how development priorities must be carefully justified and how it is not feasible to address all improvement needs simultaneously. Instead, the case company must prioritize initiatives that provide the greatest business value.

Taken together, these practical issues illustrate the challenges of collecting accurate, timely, and comprehensive data in a complex production environment. While the company is actively developing solutions (such as product-level traceability and improved human labor tracking), the current technological and practical limitations still reduce the reliability of metrics and constrain their role in supporting proactive decision making. Combined with the broader weaknesses discussed in the previous categories, these findings highlight the need for a more refined measurement system that can better align with the company's operational realities and strategic objectives. The following sections therefore explore concrete improvement opportunities for developing such a system, while also taking into account the current constraints and development priorities of the case organization.

4.2 Improvement Opportunities for Performance Measurement

The improvement opportunities presented in this chapter emerge naturally from the weaknesses identified in chapter 4.1.4, but they are also shaped by the current capabilities and priorities of the case organization. Since not all identified weaknesses can be addressed simultaneously, the focus here is placed on opportunities that are both feasible and likely to provide meaningful support for production in the near and medium term. For this reason, the topics discussed in this chapter do not correspond one-to-one with the weaknesses outlined earlier, but instead represent areas where development is expected to have the greatest impact.

After the strengths and weaknesses of the current metrics were identified, the development opportunities were discussed and refined together with representatives of the case company in two separate validation workshops. The first workshop focused on validating the empirical findings, prioritizing development directions, and refining preliminary ideas. The second workshop concentrated on presenting more elaborated development directions and exploring potential implementation steps and timelines. Details of the two workshops are summarized in table 8.

The following subchapter introduces the main focus areas for potential improvements, which were identified through the data analysis and discussed during the first workshop. These areas form the basis for the more concrete improvement proposals that are presented later in chapter 4.2.

Table 8: Details of validation workshops conducted with representatives of the case company.

Workshop number	Participants	Date	Workshop length
W1	Interviewees 1, 3, 4, and 6	10.9.2025	1:02:22
W2	Interviewees 1, 3, 4, 6, and 11	20.10.2025	54:40

4.2.1 Focus Areas for Improvement Opportunities

Before presenting the development directions, it is important to consider the strategic goals of the case company, as these are ultimately tied to the development targets of the factory organization. As discussed in chapter 4.1.2, the company's high-level strategic objectives emphasize three main factors for the Vieremä plant: (1) factory safety, (2) factory productivity, and (3) product quality.

From the perspective of the factory organization, this means that cost-effectiveness and continuous optimization of costs are not directly linked to the factory's definition of productivity. Although factory management regularly monitors indicators such as value added during the production process, these measures do not strongly influence daily operations. As emphasized by several interviewees, component costs are primarily managed by the procurement organization, and differences in the sales value of finished machines are not directly relevant to factory-level activities, which are instead driven by takt time and order backlog. The role of the factory is therefore to ensure efficient and reliable production processes rather than to optimize cost per machine.

> *"I don't know, we can't really influence sales prices or what we get out of our products from a production standpoint. If someone agrees to buy our machines, it is our job to deliver them. Some machines just generate a little more revenue. If it's a significant issue, then I think it would be worthwhile problem in product portfolio management terms."*
>
> – *Interviewee 3*

From the viewpoint of this research, this implies that relatively little focus was placed on aspects of value creation such as maximizing profitability through differences in machine sales value or lifecycle performance. Instead, the identified development directions are evaluated primarily in terms of their usefulness for supporting the operational performance of the factory.

This distinction reflects the broader separation between strategic and operational levels of performance measurement. At the strategic level, considerations such as the value created by different product models, life cycle profitability, and market competitiveness play a central role. In contrast, at the operational level the emphasis is placed more on optimizing production processes, resource allocation, and workload balance. Given that this research focuses on the factory organization, the following development directions are concentrated on the operational perspective.

This separation was also emphasized in the first validation workshop. Company representatives noted that both levels are important for the factory, but that some of the current measurement tools do not yet enable operational factors to be measured in sufficient detail to fully understand the drivers behind the high-level performance figures. This observation further guided the development directions of this research toward the refinement of metrics that support the factory's operational activities more directly. Nevertheless, several of the insights also carry implications for strategic measurement, which may serve as valuable avenues for further research.

1. The current productivity metric is not suited for factory-level operational development:

The first focus area for improvement concerns the suitability of the current productivity metric as a tool for operational development. This need was already identified in the early stages of the research, as the metric currently represents the factory's only measure of productivity, but its practical use and benefits remain limited.

At present, productivity is defined as the ratio of total working hours to the number of completed machines. While this formulation provides a high-level view of resource efficiency, it is not sufficiently precise to support continuous improvement at the factory level. As discussed in section 4.1.2, the metric is conceptually similar to the Mills Productivity Index, which was originally designed for industry-level measurement. In practice, it only reflects the relationship between human labor input and final output, while excluding other relevant inputs such as materials, capital, and energy consumption. In academic literature, such omissions have been described as a lack of "completeness" in productivity measurement (Goshu et al., 2017), which is further related to the concepts of total and partial productivity discussed in chapter 2.2.1.

Within the framework presented in figure 6, this limitation can be interpreted as a misalignment between the factory's current performance measurement system and its broader strategic orientation, as the productivity measure lacks a clear definition of what constitutes value creation in the production process. This is further reinforced by the company's own ambiguity regarding how productivity should be defined in the current context:

> *"However, I'm not entirely sure how productivity should be described in general. There have been all kinds of ideas about it, but I don't know what the best way to describe productivity really is."*
>
> *– Interviewee 10*

The interviews revealed that there is no clear consensus on how productivity should be understood within the factory. Several employees emphasized that productivity should not be assessed solely through aggregated output-to-input ratios, but should also reflect operational factors that determine the *effectiveness* of conducted work. In particular, factory personnel highlighted the importance of capturing the concept of *workload*, rather than merely recording the number of paid work hours. A metric based on workload would provide the organization with a means to evaluate the efficiency of human labor more accurately, instead of relying on lagging figures that summarize only total hours worked.

However, human work is not currently trackable in the case company's MES system, meaning that the contribution of manual labor cannot be systematically monitored in real time. In addition, the high product variation in the production process makes the evaluation of workload inherently difficult, as different machines require varying amounts of effort during assembly. One interviewee highlighted this challenge when discussing workload estimation:

> *"You would need to know which machines are difficult to make and which ones are easy. In practice, this requires experience, because when a new [person] joins the company, they really don't know. You can't tell it from the MES either, until you've been working there for 6-12 months."*
>
> *– Interviewee 5*

One possible approach to address this issue would be to further evaluate the *theoretical workload* of production processes using data already available within the company. The company already evaluates the workload of manufacturing processes through clocking. Using this information to forecast workload requirements could enable more accurate workforce planning and could, in principle, support the redistribution of labor across workstations to balance workloads and avoid bottlenecks.

It is also important to note that micro-optimization at the level of daily workforce allocation is neither feasible or desirable, as staffing levels cannot be flexibly adjusted on such short timescales. The role of an improved metric would therefore be to support medium to long-term development, such as production planning and continuous improvement activities, rather than daily resourcing decisions. The practical formulation and potential use of a theoretical workload metric are explored in more detail in chapter 4.3.

2. The performance measurement system lacks comparability:
A second improvement opportunity concerns how changes in the values of the current performance measures are interpreted by management. More precisely, these changes are not consistently compared against clear and relevant baselines over longer time periods. In several of the existing major metrics, reference levels or target values for evaluating performance are either missing or only loosely defined, which limits their interpretability. In academic literature, such omissions have been described as a lack of *comparability* in performance measurement (Goshu et al., 2017).

This issue is particularly evident in three of the factory's major metrics: the productivity metric, buffer levels, and scheduling hold. In each of these cases, the company records absolute values over time but either (1) does not systematically compare them to historical values or goal levels, or (2) compares them to targets that are not meaningful for operational development. For example, buffer sizes and scheduling hold are both monitored actively on a daily basis, but no clear threshold values exist to indicate when the levels are too low or too high. As a result, interpreting the metrics requires substantial experiential knowledge of factory operations, and the development of the metrics remains opaque to much of the organization. Defining more clear reference levels would make the performance trends of these metrics more transparent and easier to interpret across different organizational levels.

The same problem applies to metrics that already have defined goal levels. For instance, the current productivity metric includes a nominal goal level defined as the twelve-month moving average reduced by five percent. In practice, however, this value has remained largely unchanged for an extended period, and its connection to actual production capacity or process efficiency is weak. As one interviewee noted:

> *"The 'hours per machine' metric is also bad in the sense that it does not indicate the starting level. The current goal level has just been decided on a whim."*
>
> *– Interviewee 3*

From a management perspective, the absence of well-defined reference levels complicates the interpretation of results and weakens the link between operational performance and strategic goals, as the goal level does not take into account the operational realities of the company. Within the framework introduced in figure 6, this deficiency can be seen as a gap between the CI philosophy applied by the case company and the information produced by current metrics. Without comparable targets, deviations cannot be meaningfully assessed or used to guide corrective actions.

Furthermore, the company already possesses the data and technological capabilities required to establish more rigorous and meaningful baselines. For example, the target level for the productivity metric could be tied to the factory's theoretical production capacity, which would automatically account for factors such as holidays, shift structures, and production stoppages. Similarly, statistical control limits could be defined for the scheduling hold metric to identify when deviations from planned schedules are significant rather than random variation.

Developing such comparative baselines would allow the company to distinguish between normal process variation and genuine performance changes, thereby improving both daily operational control and strategic monitoring. This approach would also align the company's measurement practices more closely with the principles of continuous improvement, where measurable, evidence-based targets form the basis for learning and development.

3. The performance measurement system does not link cause and effect:
A third improvement opportunity concerns the performance measurement system's limited ability to identify cause-and-effect relationships within the production process. While the factory already monitors several key performance metrics, these indicators do not yet enable systematic analysis of the underlying factors that drive performance variation. As a result, changes in metric values are often observed without a clear understanding of their root causes.

As discussed in chapter 4.1.3, the scheduling hold metric is one of the factory's most important measures, as it aggregates information about the overall flow reliability of the production process. However, variation in scheduling hold is not currently measured or analyzed in relation to other production factors. Production occurrences such as disruptions, product variation, or new product development (NPD) projects are examples of factors that can significantly affect the plant's scheduling hold by creating process disturbances. Furthermore, deviations that occur during the parts manufacturing process or in the early stages of the assembly process often create downstream bottlenecks, but the links between these events and their cumulative effects on scheduling hold are not currently fully understood.

> *"Then we have a blind spot where, if we are missing a part, it generates other disruptions. It is currently difficult to see what causes these new disruptions, when in reality they are linked to the initial part shortage."*
>
> *– Interviewee 4*

This example illustrates how the absence of integrated data prevents the company from recognizing causal relationships between events that occur in different stages of production. Within the framework presented in figure 6, this challenge represents a gap between the innate characteristics of the manufacturing process and the information currently captured through measurement. Variability in production is a natural feature of manufacturing systems, but without systematic monitoring, it remains difficult to distinguish between random variation and variation driven by specific root causes.

Measuring and visualizing production variability would allow the company to identify and quantify the effects of different disturbances more effectively. This, in turn, would support efforts to minimize variation, which is another key principle of Lean manufacturing and CI. Additionally, integrating variability analysis with other factory-level data would provide the foundation for more advanced statistical and causal analysis. Such an approach would enable the factory to link observed outcomes (e.g. in schedule adherence or buffer fluctuations) to specific upstream events. In practice, this would make it possible to move from descriptive monitoring toward a more predictive and diagnostic performance measurement system.

Together, the three focus areas discussed above highlight the key improvement directions for the case company's performance measurement practices. They collectively emphasize the need to strengthen the diagnostic, statistical, and causal depth of performance measurement. Before introducing the concrete improvement proposals developed to address these opportunities, the following section presents key insights that emerged from triangulation interviews with representatives of other OEM companies.

4.2.2 Findings From Other OEM Companies

To broaden the perspective of the study and validate the findings derived from the case company, triangulation interviews were conducted with representatives of two other OEM manufacturers participating in the TwinFlow project: Konecranes and Mitsubishi Logisnext (later referred to as Logisnext). Both organizations operate in the field of heavy equipment manufacturing, specializing in lifting and material-handling technologies. Konecranes focuses primarily on the production of overhead cranes and industrial lifting systems, while Logisnext manufactures industrial forklifts and related logistics solutions. Due to their shared focus on assembly-intensive, high-value products, their production environments are comparable to that of the case company.

Both OEMs operate multiple production sites, including facilities located outside Finland. However, the interviews conducted for this study focused specifically on the activities of their selected Finnish manufacturing plants, as these offer the closest contextual match to the case company's production environment.

It should be noted that, due to time constraints and the scope of the external interviews, the discussions with the two OEM companies concentrated primarily on productivity-related metrics and practices. While the case company analysis in this thesis examines performance measurement from a broader perspective, the triangulation interviews were intended mainly to provide complementary insights into how comparable OEM manufacturers monitor and evaluate productivity at the factory level.

Despite differences in products and organizational scale, both companies have organized their production operations around similar core principles, including the combinations of parts manufacturing and assembly processes guided by the MTO model, and high product variation. The interviews revealed several key themes related to performance measurement and production control, which provide valuable reference points for the development of the case company's own measurement practices. An overview of the key characteristics and productivity measurement approaches of the three OEMs is presented in table 9.

Table 9: Overview of production structure and key productivity metrics among the studied OEM companies.

Company	**Structure of the selected production line(s)**	**key productivity evaluation metric(s)**
Ponsse (Case company)	Single mixed-model parts manufacturing and assembly line with some parallel subassembly cells.	**1.** Scheduling hold (compared to takt time) **2.** Working hours per finished machine
Logisnext	Multiple mixed-model parts manufacturing and assembly lines with a high-degree of parallel subassembly cells.	**1.** Scheduling hold (standard times vs. actual assembly hours)
Konecranes	Multiple mixed-model assembly lines with some parallel subassembly cells.	**1.** Scheduling hold (standard times vs. actual assembly hours) **2.** Financial productivity (OpEx/Sales)

Although the table only provides a high-level overview of the studied production systems, it still highlights notable differences in how productivity is defined and evaluated across the three companies. These differences can be partly explained by the varying nature of their products, but also by differences in production line configurations and control logic (e.g. takt time -driven versus push/pull-type production). While the detailed operational logic of each company is beyond the scope of this thesis, the following paragraphs highlight the relevant findings from the triangulation interviews and discuss their implications for the operations of the case company.

Focus on schedule adherence. The first notable takeway is related to the baseline that productivity is measured against. Interviewees emphasized the fact that their organizations currently measure productivity mainly through comparing the actual manufacturing times of their production lines to the theoretical estimates. These estimates have been calculated in advance at a selected level, as stated by one interviewee:

> *"Yes, we have calculated standard times for those products. In other words, the manufacturing time for the basic product has been calculated, and then the additional customization options are added on top of that. That's one thing we monitor, i.e., how close we can get to those pre-calculated manufacturing times. They are usually bundled into the product family level."*
>
> *– Interviewee E1*

This approach is in line with the weaknesses identified at chapter 4.1.4, as it indicates that productivity is best assessed through schedule adherence rather than through the total number of completed units at the factory level.

A key distinction compared with the case company lies in the reference framework used for productivity monitoring. In the case company, schedule adherence is evaluated relative to the takt time of the production line. While takt time can also serve as a proxy for theoretical manufacturing time, it does not capture variation between different product types or model configurations. By contrast, the use of standard times at Konecranes and Logisnext allows for monitoring deviations on a more product-specific level and enables evaluation of workload consistency despite product customization.

In addition to schedule-based indicators, Konecranes also tracks productivity from a financial perspective by comparing operational expenditure (OpEx) to sales revenue. This metric provides a higher-level view of overall cost efficiency:

> *"At the factory level, we measure operating costs as a percentage of sales. This can be monitored at the unit level also."*
>
> *– Interviewee E2*

However, while such financial metrics offer valuable long-term insight into business-level productivity trends, they are less applicable for short-term operational control and continuous improvement at the factory level. Consequently, schedule adherence remains the most practical and actionable productivity indicator for daily production management. However, the interviewees from Konecranes also mentioned how measuring the realized working hours can be misleading as a standalone indicator, since they exclude time spent on supporting or developmental activities, such as training and continuous improvement efforts.

Interviewees also noted that measuring productivity at a highly granular level remains challenging. Monitoring the conducted work in real-time would require additional technological solutions throughout the production lines, which may affect the workflow of employees. Also, additional regulative factors such as the European Union's General Data Protection Regulation (GDPR) affect the collection and processing of personal data, further complicating granular measurement.

Employee goals and rewarding. Another important takeaway concerns how productivity metrics are used to guide employee work and link performance outcomes to rewarding mechanisms. Both interviewed OEMs apply their productivity metrics not only for managerial monitoring but also as tools for communicating daily production expectations and aligning employee behavior with operational targets.

At Logisnext, factory productivity is closely connected to short-term production goals. Through production planning, employees receive clear daily output targets, which form the main reference point for line operators when tracking progress. As one interviewee explained:

> *"[Productivity metrics] are evaluated more at the management level. Line workers focus more on daily production goals and results of previous workdays. They are used more frequently in communication."*
>
> *– Interviewee E1*

These daily production goals are partly enabled by the shorter takt times of Logisnext's assembly lines. This setup ensures that operators maintain continuous visibility of production progress and provides a rapid feedback mechanism between planned goals and execution.

At Konecranes, the connection between productivity and employee rewarding is even more direct. The plant's productivity metric has a measurable impact on employee compensation, forming a significant part of the monthly bonus structure. This creates a strong link between individual productivity and pay, reinforcing personal accountability and incentivizing efficient work performance:

> *"Another factor related to productivity is performance-based pay. We calculate a direct productivity figure that affects employees' pay. [...] However, this cannot be used directly as a measure of factory productivity, as it provides so little insight into the overall picture."*
>
> *– Interviewee E3*

A somewhat similar model also exists at the case company, where a portion of monthly bonuses is tied to factors such as scheduling hold and safety performance. However, when compared to Konecranes, the linkage between measured productivity and individual compensation is less explicit. The case company's takt time -based production model emphasizes line-level scheduling hold rather than individual efficiency. While this facilitates a stable and synchronized production flow, it may also obscure individual contributions to productivity. Consequently, the incentive to improve personal efficiency is weaker compared to systems where output and performance can be traced and rewarded at the individual level.

In summary, the triangulation interviews suggest that the design of productivity metrics plays a central role not only in operational monitoring but also in shaping employee motivation and behavior. The findings further support the observation that the case company's takt time -oriented approach emphasizes stability and coordination over granular monitoring and performance of individual workers. The next chapter builds on these insights to formulate concrete improvement proposals for the case company's performance measurement system.

4.3 Introduction and Evaluation of Proposed Improvements For Current Performance Metrics

Building on the weaknesses and improvement opportunities identified in the previous sections, this chapter presents three proof-of-concept (POC) –level development suggestions for enhancing the case company's performance measurement practices. Each proposal illustrates in concrete terms how the company could leverage its existing data, systems, and operational capabilities to strengthen its decision making capacity and improve the analytical and practical value of its performance metrics.

The suggestions focus on addressing the key challenges identified earlier. In addition to outlining the conceptual design of each proposal, this chapter evaluates their practical feasibility and discusses the data requirements and organizational factors that influence their implementation in the current factory environment.

The development proposals and their implications were further discussed and refined during the second validation workshop with company representatives. Observations and insights emerging from this workshop are summarized later in chapter 4.3.1, which focuses on implementation feasibility and prioritization.

1. Using theoretical and actual workload in production planning

The first improvement proposal focuses on establishing a systematic method for estimating and utilizing the *theoretical workload* across the factory's production process. The goal of this approach is to provide a more accurate and comparable baseline for assessing the efficiency of human labor, which was identified as one of the key weaknesses in the current performance measurement system.

At the time of writing this thesis, workload information is occasionally collected through manual clocking of work tasks, measuring the time spent on individual phases production process. These results are then collected and averaged to assess workstation-specific workloads relative to the takt time. While this information is valuable for retrospective analysis, it could also be further utilized in production planning. To improve the factory's ability to anticipate resource needs and balance workloads, the company could begin estimating the expected workload for different workstations in advance.

In principle, evaluating the theoretical workload (WL^{theor}) could already be achieved by combining existing production data with additional time-based estimates. The required input data includes:

- H = planning horizon (e.g., one shift, workday, or week)
- V = set of product models planned to be built during H
- D_v = planned number of units of product variant $v \in V$ during H
- S = set of selected workstations

These variables represent information that is already readily available to the company. However, additional data points are required for accurate workload estimation. Critically, the company would first need to define the **standard time** ($T_{v,s}^{stand.}$) required to assemble each product model v at each workstation s. These standard times could be derived from the existing clocking data already collected by the company, and they could be further strengthened by including the effects of different customization options.

The increased use of such standard times in production planning would align the case company's practices more closely with those of the other OEMs described in Chapter 4.2.2. Once the required parameters are defined, the theoretical workload can be calculated as follows:

Total theoretical workload at a single workstation s during H:

$$WL_s^{theor.} = \sum_{v \in V} D_v T_{v,s}^{stand.} \tag{8}$$

These individual workload estimates can be further aggregated to evaluate the expected total workload across multiple workstations:

Total theoretical workload at workstations $s \in S$ during H:

$$\sum_{s \in S} WL_s^{theor.} \tag{9}$$

The estimation of theoretical workload is particularly valuable because of its flexibility. It can be calculated for different time horizons and workstations, supporting multiple managerial purposes such as production planning and bottleneck identification. The potential applications for the resulting workload estimates primarily relate to strengthening operational decision making. Some initial use cases include:

- A new factory-level metric representing "theoretical work hours per finished machine", allowing the comparison of planned workload with realized output and improving the accuracy of productivity assessments compared to current "working hours" figures.

- Improved workstation-level forecasting, enabling better estimation of required personnel and balancing workloads across stations.
- A quantitative baseline for evaluating the cost implications of outsourcing or insourcing specific production steps, as changes in workload can be translated into labor cost impacts.

In practice, this approach builds upon the company's existing time-tracking data and would primarily require broader and more systematic use of those datasets. As a future development step, the company could also begin capturing more comprehensive real-time data from the MES. Accurate timestamp data from MES would make it possible to track the *actual workload* (WL_s^{actual}) beyond manually clocked operations, reflecting the true time distribution across the production process.

If actual workload data becomes available, comparing the theoretical and actual workloads would allow the company to measure workstation-level efficiency as follows:

$$\text{Workload Efficiency (\%)} = \frac{WL_s^{theor.}}{WL_s^{actual}} \cdot 100 \tag{10}$$

While theoretical workload could become a valuable tool for the case company, its reliability depends heavily on the accuracy and representativeness of the standard time data. Another important consideration is product variability, as extensive customization may distort standard time estimates. The accuracy of MES timestamp data is also a critical factor, since incomplete or inconsistent data collection may lead to misleading efficiency values. Additionally, to ensure correct interpretation of workload metrics, a clear understanding of employee allocation across stations is essential, particularly in cases where operators rotate between tasks or are responsible for multiple work areas during a shift.

Once implemented, this approach would enable the factory to move beyond aggregate "hours per machine" figures and toward a more granular, data-driven understanding of how labor time is distributed and how production efficiency varies across product models and workstations.

2. Linking productivity target levels to theoretical capacity

The second improvement proposal focuses on refining the factory's existing productivity metric by linking its target level to the theoretical production capacity of the plant. Whereas the first proposal introduced a new measurement perspective based on theoretical workload, this concept aims to enhance the interpretability and relevance of an already established metric.

Currently, the factory-level productivity metric is compared against a fixed goal level derived from the twelve-month moving average reduced by five percent. As discussed in chapter 4.2.1, this static reference value does not adequately account for the actual production conditions of each period. Consequently, the metric provides limited insight into whether observed changes reflect genuine performance variation or simply differences in operational circumstances.

To improve comparability, the factory could adopt a dynamic target level based on its *theoretical output capacity*. This approach anchors the productivity target to the number of machines that could theoretically be assembled during a given month, given the available working time, the factory's takt time, and the target workload for production. An illustrative comparison between the current and proposed approaches is presented in figure 12.

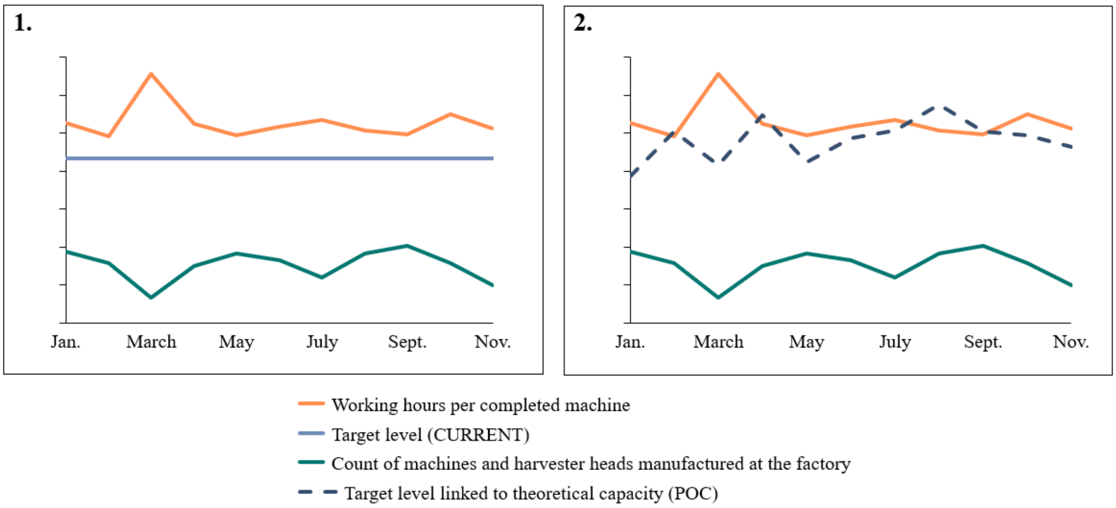

Figure 12: Illustrative comparison between productivity target levels currently used by the case company (1) and the proposed approach (2).

The figure illustrates both the current used and the proposed method for productivity target level. Graph **1** represents a mock-up of the actual productivity chart currently used by the case company. It includes some of the same key variables that are presently monitored, but the underlying data has been masked to protect confidentiality. The pattern of variation resembles actual monthly productivity data, while absolute values and scaling have been modified and are hidden from view.

Graph **2** illustrates the logic of the proposed POC productivity target. It mirrors the current graph in most respects, except that the fixed goal level has been replaced with a dynamic target based on the factory's theoretical production capacity. In this suggestion, the theoretical output for a given month is calculated using available temporal information. Specifically, the number of working days, the daily working hours, and the defined takt time of the assembly process:

$$\text{Theoretical Output} = \frac{\text{Working Days per Month} \cdot \text{Working Hours per Day}}{\text{Takt Time}} \quad (11)$$

However, this theoretical output needs to be adjusted depending on the target workload that the factory aims for, as is does not account for losses that are caused by non-value-adding work time. This adjusted realistic output can then be used to establish a monthly dynamic target level for the productivity metric:

$$\text{Target Productivity Level (POC)} = \frac{\text{Actual Working Hours in Production}}{\text{Theoretical Output} \cdot \text{Target workload (\%)}} \tag{12}$$

This relatively simple adjustment enables the company to express monthly productivity relative to the theoretical number of machines that could be produced under ideal workload level. The resulting target level dynamically reflects variations in working days and extraordinary events such as holidays or strikes. These factors are not captured by the current static target level.

In conceptual terms, this modification provides an alternative way of expressing the efficiency of working hours relative to the plant's theoretical capacity. It is closely related to the logic of the OTE metric (Muthiah and Huang, 2007), introduced earlier in chapter 2.3.2, which also evaluates performance against the theoretical potential of the production system.

Linking productivity targets to theoretical capacity offers several benefits. Most importantly, it provides the company with a more realistic view of efficiency development across months and years, as the dynamic target can automatically reflect changes in workdays, workload level or takt time. While the proposed adjustment does not directly measure true workload or station-level performance, it offers a more accurate representation of aggregate efficiency at the factory level. This supports management's ability to evaluate how effectively available labor resources are being used to meet production potential under varying operational conditions.

Implementing this approach requires only limited additional data inputs, most of which are already available in the factory's existing reporting systems. As such, the proposal represents a low-complexity improvement that can be integrated into current reporting practices without significant system modifications.

Although the proposed method improves comparability, it does not account for variations in workload or product complexity. Therefore, it should be complemented by the theoretical workload framework presented in proposal 1, which provides deeper insight into how work content and product mix influence productivity outcomes. Together, these two proposals create a more context-sensitive foundation for factory-level productivity measurement by enhancing both strategic monitoring and decision making capabilities.

3. Integrating variability and causal analysis into performance measurement

The third improvement proposal does not introduce a new metric as such, but rather focuses on enhancing the factory's ability to identify and interpret the underlying causes of performance variation by utilizing the data already available in the existing metrics. In particular, this proposal highlights the potential of analyzing variability using existing indicators (especially the scheduling hold metric) to better understand causal relationships and process bottlenecks across the production line.

Currently, scheduling hold is an important but primarily real-time measure used for identifying immediate problem areas in production. To gain a more systematic view of recurring issues over longer time periods, the factory could analyze and visualize the variability of scheduling hold data. Although this data is already recorded in the MES system, it is not yet used for trend analysis or variability tracking.

As illustrated in figure 13, the variation of scheduling hold can be visualized at the level of individual workstations using box plots, which divide the data into quartiles to illustrate its spread and skewness. These plots could, for example, display weekly variation in schedule adherence across the entire assembly line. Each box represents the distribution of workstation-level deviations from planned completion time, enabling production management to identify where variation is highest and where improvement actions should be prioritized.

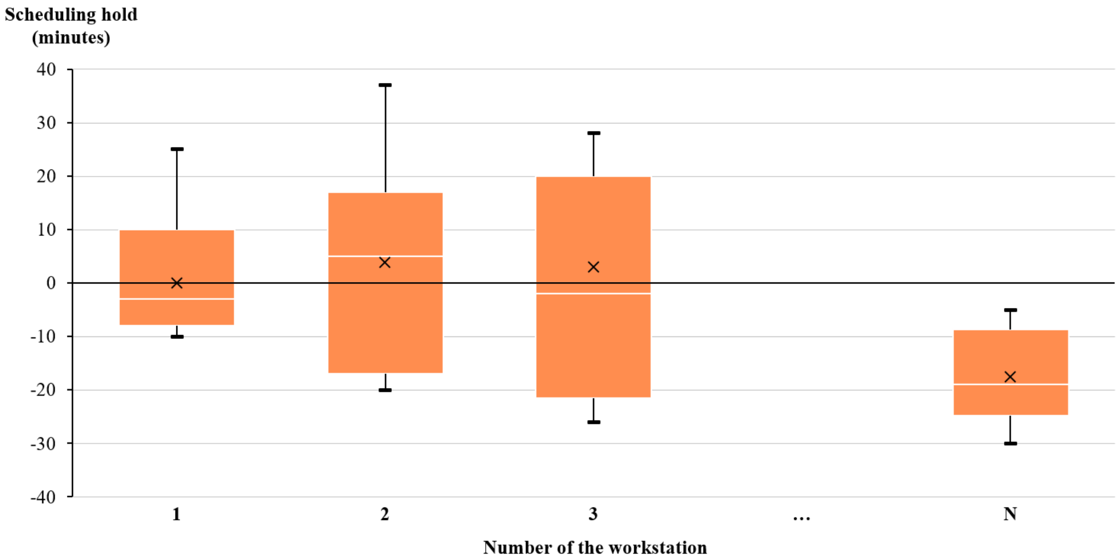

Figure 13: Illustrative display of weekly workstation-level scheduling hold box plots.

Again, the data visualized in the figure consists of simulated values that aim only to illustrate the logic of the approach. This type of visualization transforms scheduling hold from a real-time control indicator into a lagging diagnostic tool that displays more granular operational data. The proposal aims to supports a shift from aggregate factory-level performance tracking toward station-level insight.

The analytical potential of this approach can be further enhanced by enriching the scheduling hold data with complementary variables. Examples of such data include:

- **The number of disruptions** recorded at each workstation. These disruptions systematically cause delays in different phases of both parts manufacturing and assembly processes. Combining these with scheduling hold data would enable quantification of how disruptions contribute to schedule deviations across production stages.

- **The number of NPD projects** or prototype units assembled at the workstation during the same period. Based on interviews, these activities are common sources of confusion and delay, particularly in assembly operations. Similar to disruptions, combining this data with scheduling hold could help to uncover the delaying effects of these projects.

Integrating these data layers would make it possible to uncover relationships that currently remain implicit: e.g. whether delays are primarily driven by part shortages, design modifications, or workforce allocation. Moreover, this approach would externalize the "tacit knowledge" currently held by production workers and supervisors, transforming scheduling hold information into structured reports about process variability.

In the longer term, such integration could provide a foundation for more advanced analytical methods, including the use of generative AI or machine learning for pattern recognition. For instance, natural language processing tools could be employed to analyze the existing data on production disruptions. These tools could then automatically identify recurring components or defect types in the disruption logs and link them to downstream impacts in the production flow.

However, implementing this concept would require systematic integration of currently separate data sources. Although both scheduling hold and disruption data are already collected through the MES, they are not yet analyzed jointly. Furthermore, ensuring accurate and consistent data entry by production personnel is critical for maintaining data reliability.

Overall, this proposal emphasizes the importance of complementing existing performance metrics with analytical methods that capture process variability and causality. The following chapter further discusses the implementation feasibility of this and the other two improvement proposals, based on the insights gathered during the validation workshops with representatives of the case company.

4.3.1 Priorities for Implementation

The discussions held during workshops W1 and W2 also included assessments of the development priorities related to performance measurement within the case organization. These discussions covered both the practical feasibility and the expected impact of the proposed improvement initiatives, as well as the broader strategic goals of the factory regarding future performance measurement.

At a high level, the factory management expressed interest in achieving better visibility into workstation-level workload and resource utilization. Balancing workloads across the production line was identified as a particularly high priority, as the factory's MTO operating model and single-line production structure rely heavily on maintaining a stable and predictable production flow. While the factory already employs several high-level performance metrics, its visibility into detailed productivity and workload variation remains limited, making this area the primary focus of future development.

From the company's perspective, the practical implementation of the proposed improvement ideas varies notably in both complexity and anticipated benefit. The first proposal, concerning the use of theoretical and actual workload in production planning, was recognized as an interesting opportunity but was also noted to require further clarification of its practical implementation logic. While the company could already use estimates of *theoretical workload* in production planning, the limited availability of accurate *actual workload* data constrains the implementation of the solution. Therefore, the added value of this proposal remains uncertain in the short term. However, it was acknowledged that, if developed further, the proposal offers the clearest path toward a deeper understanding of workload distribution and production efficiency by building on the company's existing time-tracking practices.

The second proposal was considered the most straightforward to implement. Although its immediate benefits remain somewhat unclear, participants viewed the inclusion of more sophisticated productivity targets as a valuable addition to the current performance metric. However, the logic and true impact of the proposal need to be validated through practical implementation and comparison with actual production data over the coming months. Adjusting the target level of the current production metric would not directly alter day-to-day operations, but it could bring the measurement system closer to supporting operational-level decision making in the future.

The third proposal was the most difficult to evaluate due to its less-defined use cases. Although scheduling hold data can already be extracted from existing IT-systems, its long-term managerial relevance and integration into continuous improvement processes are not yet fully understood. Participants also noted that linking scheduling hold data to other phenomena, such as ongoing NPD projects, remains a practical challenge.

It was also acknowledged that while some improvement ideas are simpler to implement, they may yield limited functional benefits. For example, adjusting the target levels of the current productivity metric represents a relatively quick fix, but its practical value depends on long-term validation to determine whether the newly defined theoretical capacity serves as a realistic benchmark. Conversely, implementing more advanced methods to capture process variation and causal relationships between workload fluctuations and performance outcomes is likely to be more challenging, as current visibility into these phenomena remains limited.

Beyond measurement-related initiatives, the discussions also explored other methods that could support factory-level performance improvement. One recurring theme was the need for greater flexibility in reallocating operators between workstations based on real-time workload conditions. The ability to dynamically balance workloads across the line was viewed as an important future capability, particularly in an environment characterized by product variation and fluctuating assembly times. This need was also recognized during the initial interviews, as one interviewee described a potential solution for coordinating such real-time optimization:

> *"I would like to see our manufacturing line include traffic lights of some kind. In other words, if some workstation starts to fall behind, the station would show a red light. And then other workers whose workstations are currently displaying green would move to the red station to help"*
>
> *– Interviewee 9*

While this type of micro-optimization could increase responsiveness and balance short-term workload peaks, it also involves certain risks. In some cases, excessive focus on short-term adjustments may disrupt established workflows or increase process variability if not properly coordinated with broader production control principles.

Furthermore, the implementation of these practices would require that employees possess high cross-functionality, enabling reallocation across workstations.

Another important consideration raised during the discussions concerns the extent of data collection. Although more detailed information can improve the understanding of production performance, excessive or poorly designed data entry requirements may unintentionally hinder productivity or reduce employee motivation. This concern was also highlighted during the external interviews:

> *"The thing is, we need to make work time collection as easy as possible. It is always best to minimize the burden on the employee, because of course the value of the employee comes from doing the manufacturing work, not from clicking screens."*
>
> *– Interviewee E1*

This observation illustrates that changes to measurement systems can also have unintended consequences at the operational level. While more granular metrics may improve analytical accuracy, they should be designed in a way that supports the natural workflow of production employees.

Overall, the workshop discussions suggest that implementation priorities should initially focus on improving visibility into workload distribution and balancing within the constraints of the current data infrastructure. More advanced initiatives can be developed incrementally as data accuracy and system integrations improve. This phased approach allows the organization to capture early benefits from enhanced workload transparency while building the foundation for a more comprehensive, data-driven performance management system in the future.

5 Discussion and Conclusions

This thesis was conducted in response to an identified need to better understand the drivers of company performance from a metrics-based perspective. In addition to formulating specific improvement opportunities for the case company, the research also provided a broader examination of how performance is conceptualized and measured in discrete manufacturing environments. By combining theoretical analysis with empirical investigation, the study aimed to bridge the gap between high-level performance frameworks and their practical application in factory-level management.

This section summarizes the key insights derived from both the empirical case analysis and the reviewed literature, with the goal of answering the study's research question and elaborating on its theoretical implications. The first subchapter discusses the main findings of the research in relation to the study's research questions. The second subchapter connects these findings to the theoretical framework introduced in chapter 2.5, elaborating on how the results contribute to the understanding of productivity measurement in manufacturing. The third subchapter addresses the methodological and contextual limitations of the study, including considerations related to generalizability, scope and data collection. Finally, the fourth subchapter outlines potential direction for future research and practical development.

5.1 Interpretation of Results

The results of both the conducted literature review and the empirical case study provide comprehensive answers to the sub-questions of this thesis, which in are used as the basis for addressing the main research question. Through these sub-questions, the study aimed to understand the role of performance measurement in the case company's operations, while also highlighting key differences between the theoretical perspectives presented in the literature and the practical insights derived from the case analysis.

The first sub-question, *"What are the strengths and limitations of commonly used performance metrics in manufacturing?"*, was addressed primarily through the literature review. The review revealed that a wide range of well-established models exist to evaluate manufacturing performance from different dimensions, including productivity, quality, flexibility, time and cost efficiency (Neely et al.,1995; Hon,2005). Widely applied measurement frameworks such as throughput analysis, Overall Equipment Effectiveness, and Overall Throughput Effectiveness offer structured ways to measure operational performance. The main strength of these metrics lies in their ability to provide quantifiable, comparable indicators that link operational activities to higher-level business objectives.

However, many of the established metrics struggle to capture the increasing complexity of modern manufacturing systems, particularly in environments characterized by high product variety and multiple process steps. Furthermore, performance metrics are usually not capable of addressing multiple dimensions of performance simultaneously. As a result, manufacturing firms often construct their own combinations of indicators to form a performance measurement system. This process of measure integration and prioritization is inherently context-dependent and requires careful balancing between interpretability, managerial usability and implementation budget.

The second sub-question, *"What are the key improvement areas in the performance measurement practices of the case company, and how can metrics be aligned with them?"*, was explored in detail in chapters 4.1.4 and 4.2. The case company already employs a comprehensive set of metrics to monitor its operations from multiple performance dimensions. However, the analysis revealed that its current system offers limited visibility into factors such as workstation-level efficiency and resource utilization. The main improvement needs therefore concern the operational level of performance management, particularly the ability to use metric data for supporting day-to-day decision making.

The improvement proposals presented in chapter 4.3 aimed to bridge this gap by introducing approaches that better connect actual operational performance with existing measurement practices. By incorporating workload-based indicators, changing existing goal levels, and increasing the use of scheduling hold data, the company could achieve a more granular understanding of how resources are utilized and where variation occurs. Such alignment would allow management to translate performance data more effectively into actionable operational insights.

The third sub-question, *"What types of data and system capabilities are required to enable more accurate and relevant performance measurement?"*, was analyzed through both the literature review and empirical findings. Recent developments related to the Industry 4.0 paradigm offer significant potential in enhancing performance measurement through solutions that enable the collection of more accurate and real-time production data. In practice, this requires the use of modern technologies (e.g. IIoT) and carefully designed IT-systems.

However, the case study revealed that collecting and aggregating accurate operational data remains a challenge in practice. Limitations in data reliability and system integrations often hinder the full utilization of measurement potential. Particularly at the workstation level, data collection processes may still rely on manual inputs, which can harm data quality or introduce bias to the collected data. Therefore, while the technological foundations for advanced performance measurement are increasingly available, their effective use depends on organizational readiness and the careful design of measurement systems that do not burden employees or disrupt production flow.

Building on these findings, answering the main research question *" In what ways can factory-level performance metrics be refined to enable more effective decision making and contribute to sustained productivity improvement in discrete manufacturing environments?"* requires considering the entire performance measurement ecosystem that a manufacturing company utilizes. This question is inherently case-specific, as the measurement needs of manufacturing organizations are shaped by their distinct product portfolios, production structures, and strategic priorities. Therefore, rather than identifying individual improvement suggestions, the question can be better answered by identifying the key structural elements that form an effective performance measurement ecosystem.

The results of this study suggest that effective factory-level performance measurement system must balance two complementary objectives:

– On one hand, metrics should provide sufficient operational granularity to guide daily decision regarding factors such as resource allocation and workload balancing.

– On the other hand, they must remain connected to the strategic goals of the organization, enabling high-level monitoring and long-term development.

This duality highlights the importance of interplay between multiple performance metrics. As visualized in figure 3, a single measure is rarely sufficient to capture multiple different performance factors. The same limitation also applies to strategic and operative objectives, as the decision making processes in these two levels require different information. Therefore, manufacturing organizations must design a performance measurement system that integrates a diverse set of metrics addressing different temporal and hierarchical levels of decision making. These metrics should be tailored to the company's specific operational context while maintaining comparability through clearly defined baseline or target values.

The following subchapter elaborates on this phenomenon by discussing the contributions of the research and presenting an extended conceptual framework that illustrates how the performance measurement ecosystem of a manufacturing organization should ideally be structured.

5.2 Theory Elaboration

The findings of the case study contribute to the elaboration of performance measurement theory by synthesizing fragmented concepts into a prescriptive framework for structuring an effective measurement ecosystem. The analysis supports the existence of multiple, interdependent relationships between the core concepts introduced in chapter 2, which are rarely formalized in the studied literature.

The initial conceptual framework (figure 6) established that strategic orientation and operational realities jointly drive the selection of performance metrics. The empirical case study confirmed this logic, but also revealed a critical gap: the framework did not specify how individual metrics should be utilized in practice to serve different organizational processes effectively.

Building on this foundation, the elaborated framework presented in figure 14 extends the theoretical model by illustrating how the individual metrics used by a manufacturing organization should be deliberately positioned and grouped to form different sets:

USAGE IN ORGANIZATIONAL PROCESSES / TEMPORAL ORIENTATION	TOOL FOR DAILY OPERATIVE CONTROL	MONITORING STRATEGIC OBJECTIVES
LEADING (PREDICTIVE)	**1: Metrics for steering and predicting** **Use cases:** -Short-term forecasting (e.g. labor) -Identification of potential bottlenecks **Most relevant performance factors:** Quality, Flexibility, Productivity	**2: Metrics for anticipating and developing** **Use cases:** -Estimating the effects of long-term capability development actions **Most relevant performance factors:** (All)
LAGGING (RETRO-SPECTIVE)	**3: Metrics for control and reacting** **Use cases:** -Short-term monitoring of production -Reacting to deviations/disruptions **Most relevant performance factors:** Quality, Time	**4: Metrics for reviewing and reporting** **Use cases:** -Long-term monitoring of production **Most relevant performance factors:** Productivity, Cost

Figure 14: Elaborated performance measurement framework based on the findings of the case study, categorizing measurement ecosystem into different sets of metrics.

This framework aims to combine multiple relevant performance measurement concepts into an actionable table. Authors such as Neely et al. (1995) and Hon (2005) established that performance in a manufacturing context is fundamentally multidimensional, consisting of different performance factors. Separately, the work of Melnyk and Stewart (2004) introduced the concept of sorting metrics along two critical axes: usage in organizational processes (operative vs. financial) and temporal orientation (leading vs. lagging). Critically, neither article explicitly formalized how these concepts should be used to support organizational activities in manufacturing. By combining these classifications and comparing them to the metrics used by the case company, the framework aims to clarify the functional structure of an effective performance ecosystem.

The presented framework is structured as a 2x2 matrix, creating four distinct quadrants that represent functional requirements for metrics within a manufacturing organization. In practice, each quadrant should contain a set of metrics that focus on producing information about the specified phenomenon. This need to combine metrics to improve operational visibility has long been recognized, as the works of e.g. Misterek et al. (1992) and Melnyk and Stewart (2004) note the necessity to use sets of metrics in measurement. The framework contextualizes this necessity in the context of discrete manufacturing by including concrete use cases for each quadrant.

Additionally, the framework highlights which performance dimensions are the most critical regarding these quadrants based on the findings from the case study (see figure 11):

- **Quadrant 1 (Operative-Leading):** These metrics are crucial for proactively managing short-term operations. The case study highlighted **quality** and **flexibility** as critical factors for this quadrant, as the case company actively utilized metrics related to these categories in its own processes (e.g., availability of parts, work-in-process). Additionally, the suggested improvement opportunities bridged the gap between this quadrant and **productivity** by considering the addition of theoretical (predictive) workload as beneficial for the case company.

- **Quadrant 2 (Strategic-Leading):** These metrics are used to evaluate how the strategic goals of the organization will be addressed. Due to the lack of specific measures utilized by the case company in this area, no single performance factor could be identified as especially relevant. This is further supported by the company-specificity of leading metrics, as stated by Zheng et al. (2019).

- **Quadrant 3 (Operative-Lagging):** These metrics provide immediate feedback on recently occured events.The case analysis highlighted the criticality of factors like **quality** (through disruption count) and **time** (through scheduling hold) in this quadrant, emphasizing their importance in supporting short-term monitoring.

- **Quadrant 4 (Strategic-Lagging):** These metrics aggregate historical data for high-level monitoring. The case company's heavy reliance on metrics like *working hours per completed machine* and *material expenses* confirms that **productivity** and **cost** represent relevant factors in this area.

This framework aims to answer the main research question of the study by providing a structured approach for managers seeking to refine their organization's measurement systems. By better understanding the relationship between fundamental performance factors and their optimal functional use cases, a measurement ecosystem can be structured to include relevant, actionable information about the production process. However, it is important to note that due to low count of metrics analyzed in the case study, the performance factor dimension in each of the four quadrants may require further refinement if additional relevant metrics arise.

As discussed in chapter 2.5, existing literature introduces several measurement frameworks that aim to address multiple performance dimensions in a balanced manner (e.g., Kaplan and Norton, 1992). However, this study argues that the emphasis between multiple performance metrics is strongly shaped by the internal context of each manufacturing organization. While maintaining visibility into all performance dimensions is essential, the weight of operational and strategic metrics must be adjusted according to the company's situational needs. Through this framework, this thesis advocates that manufacturing organizations must form clear metric groups, each targeting one of the four functional dimensions, with the specific emphasis depending on the company's current development needs.

5.3 Limitations

While eloborating on reviewed academic literature and providing practical implications for the measurement practices of the case company, this study is subject to several limitations related to the adopted methodological design, the empirical scope of the study, and potential research bias.

First, having adopted a single-case study approach, the findings of the study are inherently restricted in their applicability outside of the discrete manufacturing sector. While the design enabled an in-depth exploration of performance measurement practices in their real operational context, it restricts the statistical generalizability (Yin, 2014) of the findings. The findings should therefore be interpreted only as analytically generalizable, being potentially transferable to other discrete manufacturing environments that share similar characteristics as the studied case company and the interviewed OEMs. These characteristics include the connected line flow stucture of the production process, moderate product variety, and comparable levels of technological maturity. Additionally, as the empirical focus was limited to Finnish manufacturing plants, geographical differences may also influence the generalizability of the results. Although such differences are unlikely to significantly affect core production processes, variations in factors such as legislation or data privacy regulations could influence the collection and use of individual-level data in other contexts.

Second, the scope of the case study was deliberately limited to the company's factory organization. As a result, the study pays less attention to other organizational functions (e.g. sales, procurement, or finance), potentially limiting the emergence of alternative viewpoints regarding performance. While the chosen focus strengthens the operational depth of the analysis, it can result in the results failing to account for other potential performance relationships from a holistic company-level perspective. This limitation is further discussed in the following subchapter.

Third, as in most qualitative studies, the research is susceptible to subjectivity and potential researcher bias. The data collection and analysis were conducted largely by a single researcher, which may have influenced the interpretation of results. To mitigate these risks, several measures were taken. These include the use of the Gioia methodology ensured transparency and traceability in data analysis, incorporation of interview quotes when presenting results to reduce ambiquity, and iterative discussions with case company representatives to validate intepretations. Additionally, the collected interview data was cross-referenced with data from other complementary sources throughout the study to ensure a comprehensive understanding of the studied phenomena.

5.4 Future Research and Implementation Directions

As performance measurement and productivity improvement in industry context remain both academically and societally relevant topics, the findings of this study open several potential directions for future research and practical implementation.

The first future research avenue is related to the managerial and organizational aspects of metric-driven management. While this thesis primarily examined the role of factory-level performance metrics, less attention was given to how these metrics are actively used in decision making processes. This is exemplified by the fact that the formulated theory-elaborating model focused on how performance measurement ecosystems should be structured, without considering in detail how they are applied in practice. Further research could therefore explore the behavioral and managerial implications of performance measurement, such as how metrics influence leadership practices and employee motivation. Understanding these mechanisms would help bridge the gap between measurement and managerial action, thereby increasing the practical impact of performance measurement systems.

The second direction involves expanding the scope of performance measurement to capture value creation across the entire product lifecycle. During the early phases of this research, performance measurement evaluation from this perspective was also considered. However, as the work progressed, it became evident that monitoring this type of lifecycle value tends to be beyond the factory organization's focus area. Future research could therefore focus on measuring productivity and value creation at the enterprise level, including how production performance interacts with the sales value of manufactured products or after-sales services. Such studies could provide useful insights for portfolio management, strategic resource allocation, and assessing the business impact of development initiatives.

Finally, this study examined a company with an already established performance measurement system, which limited the emphasis placed on evaluating the maturity and incremental value of individual metrics. For organizations with less developed systems, future work could investigate the concept of "minimum viable metrics", the essential set of indicators required to maintain meaningful visibility into operations while avoiding excessive measurement complexity. Understanding how companies evolve their measurement systems over time, and how metric maturity relates to competitiveness, would provide valuable guidance for firms developing or refining their own measurement architectures.

Collectively, these research directions highlight the need to link performance measurement more explicitly to broader questions of organizational competitiveness. Addressing these themes would further advance both the theoretical understanding and practical implementation of performance measurement in discrete manufacturing contexts.

References

Adams, M., Schroer, B. J., Gunter, L., and Schroer, B. J. (2001). Impact of process variability on lean manufacturing systems. In *Proceedings of the Huntsville Simulation Conference*, San Diego, CA. Available from: https://www.researchgate.net/publication/237123012.

Almeida, A. and Cunha, J. (2017). The implementation of an activity-based costing (abc) system in a manufacturing company. *Procedia manufacturing*, 13:932–939. DOI: 10.1016/j.promfg.2017.09.162.

Birks, M. and Mills, J. (2012). *Grounded theory: A practical guide*. SAGE Publications. ISBN: 978-147-39170-6-4.

Brudan, A. (2010). Rediscovering performance management: systems, learning and integration. *Measuring Business Excellence*, 14(1):109–123. DOI: 10.1108/13683041011027490.

Cai, H., Wan, J., and Chen, B. (2024). Digital twin-driven multi-factor production capacity prediction for discrete manufacturing workshop. *Applied Sciences*, 14(7):3119. DOI: 10.3390/app14073119.

Cua, K. O., McKone, K. E., and Schroeder, R. G. (2001). Relationships between implementation of tqm, jit, and tpm and manufacturing performance. *Journal of operations management*, 19(6):675–694. DOI: 10.1016/S0272-6963(01)00066-3.

De Toni, A. and Tonchia, S. (2002). New production models: a strategic view. *International Journal of Production Research*, 40(18):4721–4741. DOI: 10.1080/00207540210158005.

Eccles, R. G. (1991). The performance measurement manifesto. *Harvard business review*, 69(1):131–137. Available from: https://hbr.org/1991/01/the-performance-measurement-manifesto.

Eisenhardt, K. M. and Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of management journal*, 50(1):25–32. DOI:10.5465/AMJ.2007.24160888.

Fabri, M.and Ramalhinho, H., Oliver, M., and Muñoz, J. C. (2022). Internal logistics flow simulation: A case study in automotive industry. *Journal of simulation*, 16(2):204–216. DOI: 10.1080/17477778.2020.1781554.

Gershwin, S. B. (2000). Design and operation of manufacturing systems: the control-point policy. *IIE transactions*, 32(10):891–906. DOI: 10.1080/07408170008967448.

Gioia, D. A., Corley, K. G., and Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the gioia methodology. *Organizational research methods*, 16(1):15–31. DOI: 10.1177/1094428112452151.

Gomes, C. F., Yasin, M. M., and Lisboa, J. V. (2004). A literature review of manufacturing performance measures and measurement in an organizational context: a framework and direction for future research. *Journal of Manufacturing technology management*, 15(6):511–530. DOI: 10.1108/17410380410547906.

Goshu, Y. Y., Matebu, A., and Kitaw, D. (2017). Development of productivity measurement and analysis framework for manufacturing companies. *Journal of Optimization in Industrial Engineering*, 10(22):1–13. DOI: 10.22094/joie.2017.274.

Grünberg, T. (2003). A review of improvement methods in manufacturing operations. *Work study*, 52(2):89–93. DOI: 10.1108/00438020310462890.

Hayes, R. H. and Wheelwright, S. C. (1979). Link manufacturing process and product life cycles. *Harward Business Review (January-February)*, pages 133–140. Available from: https://hbsp.harvard.edu/product/79107-PDF-ENG.

Helu, M., Sobel, W., Nelaturi, S., Waddell, R., and Hibbard, S. (2020). Industry review of distributed production in discrete manufacturing. *Journal of manufacturing science and engineering*, 142(11):1–13. DOI: 10.1115/1.4046988.

Henri, J. F. (2006). Management control systems and strategy: A resource-based perspective. *Accounting, organizations and society*, 31(6):529–558. DOI: 10.1016/j.aos.2005.07.001.

Hon, K. K. B. (2005). Performance and evaluation of manufacturing systems. *CIRP annals*, 54(2):139–154. DOI: 10.1016/S0007-8506(07)60023-7.

Hopp, W. J. and Spearman, M. L. (2011). *Factory physics*. Waveland Press, 3rd edition. ISBN: 978-157-76673-9-1.

Huang, S. H., Dismukes, J. P., Shi, J., Su, Q. I., Razzak, M. A., Bodhale, R., and Robinson, D. E. (2003). Manufacturing productivity improvement using effectiveness metrics and simulation analysis. *International journal of production research*, 41(3):513–527.

ISO 22400-2:2014 (2014). Automation systems and integration — key performance indicators (kpis) for manufacturing operations management. Standard, International Organization for Standardization, Geneva, CH.

Jasti, N. V. K. and Kodali, R. (2015). Lean production: literature review and trends. *International journal of production research*, 53(3):867–885. DOI: 10.1080/00207543.2014.937508.

Kagermann, H., Helbig, J., Hellinger, A., and Wahlster, W. (2013). Recommendations for implementing the strategic initiative industrie 4.0: Securing the future of german manufacturing industry; final report of the industrie 4.0 working group. Technical report, Forschungsunion.

Kamble, S., Gunasekaran, A., and Dhone, N. C. (2020). Industry 4.0 and lean manufacturing practices for sustainable organisational performance in indian manufacturing companies. *International journal of production research*, 58(5):1319–1337. DOI: 10.1080/00207543.2019.1630772.

Kaplan, R. S. and Norton, D. P. (1992). The balanced scorecard: measures that drive performance. *Harvard business review*, 70(1):71–79. Available from: https://hbr.org/1992/01/the-balanced-scorecard-measures-that-drive-performance-2.

Ketokivi, M. and Choi, T. (2014). Renaissance of case research as a scientific method. *Journal of operations management,*, 32(5):232–240. DOI: 10.1016/j.jom.2014.03.004.

Kober, R., Ng, J., and Paul, B. J. (2007). The interrelationship between management control mechanisms and strategy. *Management accounting research*, 18(4):425–452. DOI: 10.1016/j.mar.2007.01.002.

Li, J., Blumenfeld, D. E., Huang, N., and Alden, J. M. (2009). Throughput analysis of production systems: recent advances and future topics. *International Journal of Production Research*, 47(14):3823–3851. DOI: 10.1080/00207540701829752.

Litster, J. and Bogle, I. D. L. (2019). Smart process manufacturing for formulated products. *Engineering*, 5(6):1003–1009. DOI: 10.1016/j.eng.2019.02.014.

Liukkonen, M. and Tsai, T. N. (2016). Toward decentralized intelligence in manufacturing: recent trends in automatic identification of things. *Journal of Advanced Manufacturing Technology*, 87:2509–2531. DOI: 10.1007/s00170-016-8628-y.

Manuele, F. A. (2009). Leading & lagging indicators. *Professional Safety*, 54(12):28–33. Available from: https://www.jstor.org/stable/48687557.

Mehrabi, M. G., Ulsoy, A. G., and Koren, Y. (2000). Reconfigurable manufacturing systems: Key to future manufacturing. *Journal of Intelligent manufacturing*, 11:403–419. DOI: 10.1023/A:1008930403506.

Melnyk, S. A. and Stewart, D. M. anf Swink, M. (2004). Metrics and performance measurement in operations management: dealing with the metrics maze. *Journal of operations management*, 22(3):209–218. DOI: 10.1016/j.jom.2004.01.004.

Misterek, S. D., Dooley, K. J., and Anderson, J. C. (1992). Productivity as a performance measure. *International journal of Operations & Production management*, 12(1):29–45. DOI: 10.1108/EUM0000000001294.

Muthiah, K. M. and Huang, S. H. (2006). A review of literature on manufacturing systems productivity measurement and improvement. *International Journal of Industrial and Systems Engineering*, 1(4):461–484. DOI: 10.1504/IJISE.2006.010387.

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

Nakajima, S. (1988). *Introduction to Total Productive Maintenance (TPM)*. Productivity Press, Cambridge. ISBN: 978-091-52992-3-2.

Neely, A., Gregory, M., and Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International journal of operations & production management*, 15(4):80–116. DOI: 10.1108/01443579510083622.

Ohno, T. (1988). *Toyota Production Systems: Beyond Large-Scale Production*. Cambridge, MA: Productivity Press. ISBN: 978-091-52991-4-0.

Ponsse (n.d.a). About ponsse. https://www.ponsse.com/en/company/about-ponsse, (accessed 18.6.2025).

Ponsse (n.d.b). Products. https://www.ponsse.com/en/products, (accessed 4.9.2025).

Ramasesh, R. V. and Jayakumar, M. D. (1991). Measurement of manufacturing flexibility: a value based approach. *Journal of Operations Management*, 10(4):446–468. DOI: 10.1016/0272-6963(91)90005-I.

Roller, M. R. and Lavrakas, P. J. (2015). *Applied qualitative research design: A total quality framework approach*. Guilford Publications. ISBN: 978-146-25157-5-2.

Roser, C., Ahrens, T., and Roser, J. (2025). Empirical analysis of the influence of the takt time of automotive assembly lines on the efficiency on these lines. *IFIP International Conference on Advances in Production Management Systems*, pages 469–479. DOI: 10.1007/978-3-032-03542-4-32.

Ruiz-de-Arbulo-Lopez, P., Fortuny-Santos, J., and Cuatrecasas-Arbós, L. (2013). Lean manufacturing: costing the value stream. *Industrial Management & Data Systems*, 113(5):647–668. DOI: 10.1108/02635571311324124.

Saunders, M., Lewis, P., and Thornhill, A. (2009). *Research methods for business students*. Pearson education. ISBN: 978-129-24027-2-7.

Saunders, M. and Tosey, P. (2013). The layers of research design. *Rapport, (Winter)*, pages 58–59. Available from: https://www.academia.edu/4107831.

Schreyer, P. and Pilat, D. (2001). Measuring productivity. *Economic studies*, 33(2):127–170. DOI: 10.1787/9789264194519en.

Schroeder, R. G., Bates, K. A., and Junttila, M. A. (2002). A resource-based view of manufacturing strategy and the relationship to manufacturing performance. *Strategic management journal*, 23(2):105–117. DOI: 10.1002/smj.213.

Scott, D. and Pisa, R. (1998). Can overall factory effectiveness prolong mooer's law? *Solid state technology*, 41(3):75–81. Available from: https://link.gale.com/apps/doc/A21188455/AONE?

Shahrokhi, M. and Bernard, A. (2009). A framework to develop an analysis agent for evaluating human performance in manufacturing systems. *CIRP Journal of Manufacturing Science and Technology*, 2(1):55–60. DOI: 10.1016/j.cirpj.2009.08.002.

Silvestri, L., Gallo, T., and Silvestri, C. (2022). Which tools are needed to implement lean production in an industry 4.0 environment? a literature review. *Procedia Computer Science*, 200:1766–1777. DOI: 10.1016/j.procs.2022.01.377.

Singh, H., Motwani, J., and Kumar, A. (2000). A review and analysis of the state-of-the-art research on productivity measurement. *Industrial Management & Data Systems*, 100(5):234–241. DOI: 10.1108/02635570010335271.

Taninecz, G. (2004). Faster but not better: improved manufacturing times don't always result in comparable cost reductions. *Industry Week*, 253(4):47–50. Available from: https://go.gale.com/ps/i.do?id=GALE

Walaski, P. (2020). The role of leading & lagging indicators in osh performance management. *Professional Safety*, 65(8):29–35. Available from: https://onepetro.org/PS/article-abstract/65/08/29/448759.

Wiendahl, H. P., Reichardt, J., and Nyhuis, P. (2015). *Handbook of Factory Planning and Design*. Springer. ISBN: 978-366-25205-3-6.

Wilkes, J. (2005). Leading and lagging practices in performance management. *Measuring Business Excellence*, 9(3). DOI: 10.1108/mbe.2005.26709caf.002.

Womack, J. P., Jones, D. T., and Ross, D. (1990). *The Machine That Changed the World*. New York: Rawson Associates. ISBN: 978-184-73705-5-6.

Xu, Y., Qamsane, Y., Puchala, S., Januszczak, A., Tilbury, D. M., and Barton, K. (2024). A data-driven approach toward a machine-and system-level performance monitoring digital twin for production lines. *Computers in Industry*, 157:1–14. DOI: 10.1016/j.compind.2024.104086.

Yin, R. K. (2014). *Case study research*. SAGE, 5th edition. ISBN: 978-145-22425-6-9.

Yuan, L., Guo, Y., Jiang, J., and Nian, L. (2013). The research on monitoring of discrete manufacturing process based on internet of things. In *2013 IEEE International Conference on Green Computing and Communications*, pages 1186–1191. IEEE. Available from: https://ieeexplore.ieee.org/document/6682219.

Zheng, L., Baron, C., Esteban, P., Xue, R., Zhang, Q., and Yang, S. (2019). Using leading indicators to improve project performance measurement. *Journal of Systems Science and Systems Engineering*, 28(5):529–554. DOI: 10.1007/s11518-019-5414-z.

A Internal interviews: Interview Guide and Questions

1. Intro
- Diplomitöiden aihealueiden esittely
- Maininta haastattelujen nauhoittamisesta ja tulosten jakamisesta (nimien anonymisointi, työrooli mainitaan lopputuloksessa) sekä datan tietoturvallisesta säilyttämisestä (Aallon palvelimet, henkilökohtaiset Aallon työkoneet) ja käsittelystä (Aallon hyväksymät työkalut ja ohjelmistot)
- Haastateltavan taustan kartoitus (työrooli, vastuut, osallistuminen tuotantoprosesseihin + osallistuminen tehtaan tietojärjestelmien käyttöön)

2. Operaatiot & Tuottavuus
- Kuinka manuaalisesti/digitaalisesti nykyistä tuotantoprosessia seurataan?
- Kuinka tehdasoperaatioiden toteutumista tai tuottavuutta mitataan tällä hetkellä vastuualuellanne?
  - Mitä mittareita on käytössä ja miten niitä tulkitaan?
  - Mitä kaikkea tuotantoprosessista tulisi mitata?
- Ovatko mittarit kaikkien työntekijöiden seurattavissa? Saavatko työntekijät palautetta niiden perusteella?
  - Osallistetaanko työntekijöitä operaatioiden ja tietojärjestelmien kehittämiseen?
- Vastaavatko nykyiset mittarit operaatioiden käytännön toteutumista tai todellista suorituskykyä?
  - Aiheuttavatko nykyiset mittarit haasteita tai sokeita pisteitä tuottavuuden seuraamisessa ja tulkitsemisessa?
- Onko olemassa epävirallisia tai henkilökohtaisia menetelmiä, joita käytetään prosessien tuottavuuden arvioimiseksi?

3. Järjestelmät & Data-arkkitehtuuri
- Mitä tietojärjestelmiä tehtaalla on käytössä?
  - Mitä niistä kohtaat työssäsi?
  - Mitkä järjestelmät ovat käytössä tuotantoprosessin eri vaiheissa?
    - Miten hyvin järjestelmät palvelevat näitä prosesseja?
  - Mitkä ovat kriittisimmät järjestelmät tuotannon ohjauksessa?
- Mitä tavoitteita järjestelmien kehityksessä on tällä hetkellä? Missä haluttaisiin olla?
- Onko järjestelmäarkkitehtuurista virallista dokumentaatiota (kartat, kaaviot)?
  - Päivitetäänkö dokumentaatiota systemaattisesti?
  - Kuka vastaa siitä?
- Miten tietoa siirretään eri järjestelmien välillä?
  - Onko tiedon siirrossa pullonkauloja?
    - Vaikuttavatko ne tuotannon etenemiseen?
- Mitä toiminnallisuuksia tai automaatioita toivoisit järjestelmiin, jotta ne palvelisivat sinua työssäsi paremmin?
  - Entä datan hyödyntämisen parantamiseksi?

B External interviews: Interview Guide and Questions

1. **Intro**
   - Diplomityön aihealueen esittely
   - Maininta haastattelujen nauhoittamisesta ja tulosten jakamisesta (nimien anonymisointi, työrooli mainitaan lopputuloksessa) sekä datan tietoturvallisesta säilyttämisestä (Aallon palvelimet, henkilökohtaiset Aallon työkoneet) ja käsittelystä (Aallon hyväksymät työkalut ja ohjelmistot)
   - Haastateltavan taustan kartoitus (työrooli, vastuut, osallistuminen tuotantoprosesseihin)
2. **Nykyiset Operaatiot**
   - Ylätason katsaus tehtaan tuotantoprosesseihin: Mitä tuotteita tehtaalla valmistetaan ja miten valmistusprosessi on organisoitu?
     - Mitä yleisiä tuotantoperiaatteita tehdas noudattaa (tahtiaika, imuohjaus yms.)?
     - Kuinka suurta tuote- ja prosessivariaatiota tehtaalla kohdataan?
     - Kuinka tehdään sisälogistiikkaprosessit on organisoitu?
3. **Tuottavuus**
   - Kuinka tehdasoperaatioiden toteutumista tai tuottavuutta mitataan tällä hetkellä vastuualueellanne?
     - Mitä metriikoita on käytössä ja miten niitä tulkitaan?
     - Miten mittareissa käytettävä data kerätään?
     - Mitä kaikkea tuotantoprosessista tulisi mitata?
   - Ovatko mittarit kaikkien työntekijöiden seurattavissa? Saavatko työntekijät palautetta niiden perusteella tai onko tuottavuus sidottu esim. bonuksiin?
     - Osallistetaanko työntekijöitä operaatioiden ja tietojärjestelmien kehittämiseen?
   - Vastaavatko nykyiset mittarit operaatioiden käytännön toteutumista tai todellista suorituskykyä?
     - Aiheuttavatko nykyiset mittarit haasteita tai sokeita pisteitä tuottavuuden seurannassa ja tulkitsemisessa?
   - Onko olemassa epävirallisia tai henkilökohtaisia menetelmiä, joita käytetään prosessien tuottavuuden arvioimiseksi?

C Descriptive table of metrics used by the case company

Table C1: Descriptions of the case company's major performance metrics.

Metric	Performance dimension	Description	Measurement logic
Working hours per completed machine	Productivity	Measures the total count of paid work hours per a completed machine over a chosen time period	$\frac{\text{Hours worked in production}}{\text{Count of completed machines}}$
Disruptions in production	Quality	Counts unplanned stoppages or breakdowns in the production process during a chosen time period. Indicates reliability of production flow	Count of reported disruptions
PPM-metric and claims count for supplier parts	Quality	Measures supplier quality by tracking defective parts received and claims to supplier claims due to part failures	$\frac{\text{Faulty parts}}{\text{1 000 000 delivered parts}}$
Work-in-process (WIP)	Quality	Tracks the number of incomplete units in the assembly process during a single point of time	Count of incomplete products
Availability of parts	Flexibility	Forecast of parts availability for the upcoming weeks/months based on supplier information	$\frac{\text{Available parts at required time}}{\text{Total planned parts demand}}$
Buffers in production	Flexibility	Measures the size of intermediate stock between process stages during a single point of time. Used for multiple buffers in different stages of production	Count of machines or subassemblies in buffer i
Scheduling hold	Time	Measures deviation between the scheduled and actual completion time of a machine for a given process stage during a single point of time	Scheduled completion time for machine i − Actual completion time for machine i
Personnel expenses	Cost	Tracks total labor-related costs for production staff over a chosen time period	$\sum(\text{Hours worked}_i \cdot \text{Hourly rate}_i)$
Material expenses	Cost	Tracks the cost of materials and components consumed in production over a chosen time period	$\sum(\text{Quantity used}_j \cdot \text{Unit cost}_j)$