CHALMERS
UNIVERSITY OF TECHNOLOGY

Integrating External Benchmarking for Machine Performance

A Case Study of Machine Performance at a Multinational Manufacturing Company

Master's thesis in Supply Chain Management and Quality and Operations Management

EBBA HÖRNFELDT
JOSEFINE NYVALL

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF INNOVATION AND R&D MANAGEMENT

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Integrating External Benchmarking for Machine Performance

A Case Study of Machine Performance at a Multinational Manufacturing Company

EBBA HÖRNFELDT
JOSEFINE NYVALL

Department of Technology Management and Economics
Division of Innovation and R&D Management
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026

Integrating External Benchmarking for Machine Performance
A Case Study of Machine Performance at a Multinational Manufacturing Company
EBBA HÖRNFELDT
JOSEFINE NYVALL

Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Gothenburg, Sweden 2026

Integrating External Benchmarking for Machine Performance
A Case Study of Machine Performance at a Multinational Manufacturing Company

EBBA HÖRNFELDT
JOSEFINE NYVALL

Department of Technology Management and Economics
Chalmers University of Technology

SUMMARY
Manufacturing companies rely on internal performance targets to evaluate machine efficiency, yet these targets are rarely validated against external industry levels, leaving it unclear whether performance ambitions reflect competitive standards. This thesis develops an external benchmarking initiative for evaluating machine performance through OEE-related KPIs, specifically speed efficiency, time efficiency, and material efficiency, by collecting external data via machine suppliers and adapting it to machine-specific operational conditions.

The study was conducted as a case study at a multinational manufacturing company operating seven European production plants, each running a diverse product portfolio on the same type of machine. Data was collected through interviews, observations, and extraction of internal performance data, and was analyzed in relation to supplier-provided industry reference levels.

The findings show that speed efficiency could be fully benchmarked externally by accounting for product mix through a scenario-based weighting approach, categorizing products into complexity levels and weighting them by their share of production run time. Time efficiency and material efficiency could not be benchmarked at a complete KPI level due to data limitations, but individual components could still be compared against supplier-provided percentages to calculate component-level indices.

A simulation-based integration approach was developed, where external benchmarking data is translated into each machine's own operational context. This allows performance gaps to be identified relative to industry reference levels. The initiative also provides a foundation for challenging internally defined performance targets with externally validated benchmarks. The developed benchmarking initiative is intended to be applicable to the manufacturing industry more broadly.

.

Keywords: benchmarking, OEE, speed efficiency, time efficiency, material efficiency, machine performance, product mix, manufacturing, performance measurement, external benchmarking

Acknowledgements

We would like to thank our case company for allowing us to conduct our study at their company and for providing us with the resources needed to carry out this research. We would also like to thank our supervisors for their guidance and helpful insights throughout the process.

Finally, a special thank you to our supervisor and examiner Hendry Raharjo at Chalmers University of Technology whose dedication, encouragement and thoughtful feedback throughout our weekly meetings pushed us to continuously improve our work.
sincerely Ebba Hörnfeldt and Josefine Nyvall

Contents

List of Figures

List of Tables

1

Introduction

This chapter introduces the background and context of this Master's Thesis, outlining the problem that motivates the study and the purpose it aims to fulfill. The research questions guiding the investigation are also presented. Together, these elements provide the foundation for understanding why this topic is relevant and worth examining.

1.1 Background

This section provides the background necessary to understand the context of the study. It introduces the current state of the global pulp and paper industry and the case company.

1.1.1 Industry Context

Global pulp and paper productivity, defined as the volume produced relative to the resources used, peaked in 2018 and has since declined at a rate of 1.6 percent per year, leaving productivity approximately 8 percent lower in 2021 than it was in 2016 (Allott et al., 2023). According to them, this trend is captured by the PaperLens Index (PLI), which draws on data from 59 major global players in the industry, supplemented by proprietary sources, to analyze how material of different grades is produced and with what level of resources.

The year 2018 marked a turning point for the pulp and paper industry. After that, companies spent more on capital, workforce, and operations but production did not grow at the same rate. Simply put, more investment was no longer leading to more output. In 2021, some companies managed to improve productivity by controlling capital and operational costs (Allott et al., 2023). This productivity decline takes place against a backdrop of low profitability, overcapacity, and weak demand across the industry, with intense price competition making it increasingly difficult for companies to grow margins (Leppävuori et al., 2026).

While the general trend points to decline, 19 of the 59 companies analyzed managed to increase their PLI scores between 2016 and 2021. These outperforming companies fall into two distinct groups. The first consists of companies with newer technology in their mills, achieving production volume growth alongside unit cost decreases by leveraging digital manufacturing technologies. The second group includes companies with older mill technology, where production volumes are stable or decreasing, yet high productivity is maintained through operational excellence and strict cost control (Allott et al., 2023).

According to Allott et al. (2023), three critical areas of focus have been identified for industry leaders to address the root causes of productivity decline: operational excellence, digital technologies, and capital productivity. Of particular relevance to this study is the first area, companies that have implemented operational excellence programs have achieved up to 10 percent improvement in Overall Equipment Effectiveness (OEE). This is achieved through improved machine reliability, preventive maintenance, SKU-specific speed targets, new standard operating procedures, and systems to measure and improve Key Performance Indicators (KPIs) over time. Leppävuori et al. (2026) report that leading companies have improved OEE by 1–2 percentage points through reduced failures and downtime. At the same time, only about one third of the time devoted to maintenance is actually spent on maintenance work, with the rest going to administrative tasks and other activities (Leppävuori et al., 2026). More broadly, leading companies are increasingly moving away from static planning and instead considering different market scenarios to better understand their competitive position (Leppävuori et al., 2026).

1.1.2 The Case Company

The case company is a large multinational corporation operating within the pulp and paper industry, with production sites across multiple continents. A central feature of its operations is that it controls multiple stages of its own production chain, from raw material sourcing through to the manufacturing of finished products, meaning that efficiency at the site level has direct implications for the broader production chain.

The company operates through several largely autonomous business areas, each carrying its own profit and loss responsibility. This means that individual production units operate with considerable independence, making consistent and comparable performance measurement between units both important and challenging.

A defining characteristic of the case company is its broad and diverse product portfolio, offering a wide range of product variants with frequent new product introductions.

This high degree of product variety means that production is characterized by frequent changes and large variation in product specifications, which introduces considerable complexity into performance measurement.

Operational efficiency and cost competitiveness are prioritized in the company's strategy. The ability to accurately measure and benchmark production efficiency is therefore critical, as it provides the foundation for identifying improvement opportunities and tracking progress over time. To measure machine performance a the company uses the KPI OEE to reflect how good their machines are performing. This KPI is broken down into three different KPIs which together in the OEE measure reflect the whole performance of how efficiency the planned time is utilized, how fast the machine runs and then the quality which is measured on a higher level and more focus on the total waste produced my the company. When measuring machine performance the company mainly focuses on measuring the time efficiency and the speed efficiency.

Performance benchmarking is widely recognized as a strategic tool for identifying improvement potential and evaluating operational competitiveness. The case company has recently conducted an internal benchmarking study, investigating and comparing performance measures within an internal setting by comparing only its own production plants. The next step the company desires is to understand how they perform in an external industry context, that is, how their performance stands in comparison to the broader industry, which includes competitors and other companies utilizing the same machines.

1.2 Problem Statement

The case company has established internal performance targets. However, these targets are defined internally and have never been validated against external industry levels. As a result, it remains unclear whether current performance ambitions reflect competitive standards, as it is unknown whether they are underperforming or overperforming. This was captured by a company representative, who stated that ”internally we always talk about how far we are from industry best practice, while at the same time we don't know how far.” This is why the case company wanted to understand how their machine performance could be evaluated against an external reference level through an external benchmarking study. However, there was no clear direction for how to conduct this study, and the company did not possess any external benchmarking data, making it more difficult to understand how they performed against the industry.

1.3 Statement of Purpose

The aim of this thesis is to develop an industry benchmarking initiative for evaluating machine performance through KPIs by collecting external industry data and adapting it to a machine-specific operational conditions in order to establish industry reference levels.

The benchmarking imitative offers several key benefits. It enables the company to compare machine performance against industry reference levels rather than solely internal historical data, thereby identifying performance gaps and areas of improvement. Furthermore, external benchmarking can challenge and refine existing performance targets, while integration of external reference values supports more transparent comparisons between machines.

1.4 Research Questions

To enable the development of an industry benchmarking initiative for evaluating machine performance, it was necessary to determine which data could be collected and which approach should be employed to compare internal and external data in a meaningful and fair way.

To enable a fair comparison between internal and external performance data, it was necessary to investigate whether all three KPIs could be externally benchmarked, leading to the following research question:

RQ1: Which KPIs can be benchmarked externally?

To determine how such a comparison can be conducted, the following research question was formulated:

RQ2: How can external benchmarking data be integrated with internal performance data?

1.5 Delimitations

This study examines seven European production plants operated by the case company, using performance data from the full year of 2025. The scope was limited to cross-plant variation at a single point in time, rather than long-term performance developments.

The analysis focused exclusively on one specific type of machine at each plant. All external benchmarking data were restricted to European production plants.

A holistic perspective was adopted, emphasizing high-level OEE-related KPIs rather than detailed technical production parameters or financial outcomes. The study aimed to identify and illustrate performance differences and did not seek to explain underlying causes at individual plants, but rather to provide a more general perspective.

In this study, OEE is often referred to as a KPI, and its components, time efficiency, speed efficiency, and material efficiency, are often referred to as KPI components.

1.6 Disposition of the Thesis

Chapter 1 (Introduction) introduces the background of the case company, the industry context, as well as the purpose of the thesis and research questions to be investigated. Chapter 2 (Theoretical Framework) presents the literature and theoretical foundations underpinning this study. Chapter 3 (Method) describes the research design, the strategy applied, and the methods used for data collection. It also addresses the study's generalizability, reliability, validity, and ethical considerations. Chapter 4 (Results) presents both qualitative and quantitative findings, supported by visualizations. Chapter 5 (Discussion) discusses the empirical findings in relation to the literature and outlines theoretical implications as well as areas for improvement. Finally, Chapter 6 (Conclusion) provides answers to the research questions and suggests directions for future research.

2

Theoretical Framework

This chapter presents the theoretical framework for this thesis, covering benchmarking, operational excellence, the concept of OEE, the House of Quality, and weighting methods to account for product variability in a production system.

2.1 Benchmarking

According to Hong et al. (2012), benchmarking as a method can be described as essential for organizations to improve their way of working and achieve performance targets by systematically comparing themselves to other companies in a specific context of choice. They also highlight that through such a comparison, based on predefined areas such as, for instance, specific processes and outcomes, the organization can identify performance gaps, evaluate their current position, and determine areas for improvement. According to them, this enables organizations to learn from best practices and to achieve advantages against competitors.

As described by Attiany (2014), the tool benchmarking generates a baseline for continuous improvement activities by, for example, providing an external focus on internal processes or functions. Attiany (2014) also pointed out the value of the tool by describing that it can provide learning opportunities by comparing a company's current state to best practices, which eventually will offer benefits such as cost savings and customer satisfaction, which will impact the company's profitability.

Building further on the applicability of the tool, Alderman and Murray (2025) described that there are different types of benchmarking aimed at different levels of an organization. They described that in the most generic way, benchmarking can be understood as a systematic and comparative process in which predefined variables are analyzed in a structured and repeatable manner.

Several different types of benchmarking models have been described in the literature, where the purpose of each step in the model is to outline what should be done when performing a benchmarking study. Anand and Kodali (2008) compared multiple models, ranging from those containing four steps to those containing 33 steps. One of the models they discussed was the Camp (1989) model, which consists of four phases: planning, analysis, integration, and action. The planning phase involves identifying benchmarking subjects and partners, as well as determining the data collection method. The analysis phase includes determining the current competitive gap and projecting future performance. The integration phase encompasses communicating findings, gaining acceptance, and establishing goals. Finally, the action phase was described to involve developing action plans, implementing them, monitoring progress, and recalibrating the benchmark. According to them, this can be viewed as a circular and continuous process in which a company repeatedly measures its performance against suitable peers, enabling ongoing performance improvement.

Alderman and Murray (2025) provided an overview of different types of benchmarking, emphasizing how organizations could approach benchmarking from multiple perspectives to gain valuable insights. They distinguished between formal and informal benchmarking. Formal benchmarking included performance benchmarking and best practice benchmarking, where the focus was either on comparing performance metrics or identifying and implementing best practices related to these metrics. Furthermore, they discussed benchmarking as also being classified based on the nature of the reference point. Alderman and Murray (2025) further noted that benchmarking can be conducted at several levels: internally within an organization, competitively against rivals, industry-wide across both competitors and non-competitors, generically across different industries, or as a global assessment.

In addition, Alderman and Murray (2025) emphasized that the context of benchmarking could vary depending on whether the focus was on a specific process, function, or performance. The distinction between these contexts lay in whether the analysis examined the system or outcome characteristics in quantifiable terms, such as, for instance, production speed. According to Chotayakul and Punyangarm (2026), credible and comparable performance measurement requires reliable benchmarks, which can be effectively estimated from equipment specifications when historical data are lacking.

Performance evaluation relies on the use of measurable indicators that can be assessed relative to internal goals and through benchmarking relative to external reference levels, creating an external target (Ahmad & Dhafr, 2002). However, they also pointed out that since such indicators represent absolute measures, evaluating them in isolation risks producing an incomplete and potentially misleading picture of the overall performance. Ahmad and Dhafr (2002) therefore highlighted the importance of applying a balanced combination of indicators that together reflect the strategic priorities of the organization, ensuring that performance is evaluated from a broader perspective that accounts for the multiple factors that influence it. They also stated that gaps between plants, identified through benchmarking and comparing against best practices, can be seen as areas of improvement.

According to Borrás et al. (2024), in the support of forward planning, target setting has been identified as a critical activity, as it guides the actions required to improve performance. They emphasized the importance of selecting appropriate reference points or comparable peer companies as a crucial step in establishing attainable targets that guide future directions. In this process, the criteria of similarity and comparability are considered essential.

Alderman and Murray (2025) highlighted the challenge of identifying a suitable reference point for effective benchmarking, which can limit comparability. According to Ofori-Kuragu (2025), it is not always possible to access best-in-class performers or find organizations willing to share their performance data, as potential benchmarking partners often have little motivation to engage when there are no mutual benefits. In response to this challenge, he argued that benchmarking frameworks should be designed so that comparisons can be made even without direct cooperation from a top-performing organization. Instead, organizations should be able to rely on alternative reference points that are realistically accessible. Furthermore, Alderman and Murray (2025) presented several challenges associated with best practice benchmarking, including limited cooperation from relevant stakeholders. In addition, they noted that without sufficient stakeholder participation, implementing improvements becomes difficult.

2.2 Porter's Five Forces

Porter (1980) argued that the intensity of competition within an industry is not a matter of luck, but is rooted in its underlying economic and technological structure. To analyze this structure, Porter developed the Five Forces framework, which identifies five fundamental competitive forces that collectively determine the profit potential of an industry. These forces are the threat of new entrants, the bargaining power of buyers, the bargaining power of suppliers, the threat of substitute products, and the rivalry among existing competitors. According to Porter (1980), the goal of competitive strategy is to find a position within the industry where these forces do the company the most good or the least harm. By understanding the sources of competitive pressure, a company can identify its strengths and weaknesses relative to the industry and determine where strategic action may yield the greatest payoff.

Building on the framework presented by Porter (1980), Chen et al. (2026) empirically demonstrated that firms which maintain a higher awareness of external environmental changes and competitive forces consistently achieve superior performance outcomes. In their study of 1,038 firms operating in China, they identified that firms combining strong competitive strategies with strategic resource allocation outperform those relying on isolated strategic approaches. Central to their findings is the concept of the ”Industry Leading Type”, which they described as firms that exemplify best-in-class practices and set the standard for others to follow. These firms excel across all strategic dimensions and serve as a benchmark for other firms in the industry.

Further empirical support for the relationship between Porter’s competitive forces and benchmarking is provided by Baird et al. (2024), who examined how the intensity of competitive forces affects organizational performance and competitive advantage among 505 US-based financial managers. Their findings demonstrated that benchmarking, as a contemporary management accounting practice, plays a central role in helping organizations manage the pressures identified by Porter’s framework. Specifically, through comparing their performance against competitors, organizations become more aware of threats from substitute products and new entrants, and are better positioned to improve their competitive standing. Baird et al. (2024) further argued that benchmarking enables organizations to evaluate and reassess their competitive position, making it a strategic tool for responding to the forces that shape industry competition.

2.3 Operational Excellence

According to Antony et al. (2026), operational excellence is a philosophy focused on continuously improving business processes. The goal is to improve productivity, effectiveness, efficiency, and overall performance. They described it as a systematic approach that involves all levels of an organization. It encourages fact-based decision-making and builds a culture of continuous improvement. The aim is to minimize waste, optimize resources, and deliver superior value to customers and stakeholders.

Antony et al. (2026) identified five key themes that have shaped operational excellence historically. These are: continuous improvement, quality and organizational culture, top management commitment, productivity and process enhancement, and either leadership or appropriate tools and techniques. They further highlighted that productivity and process enhancement is essential for maintaining competitive advantage. This includes identifying inefficiencies, streamlining processes, and improving agility.

Looking forward, Antony et al. (2026) pointed to digital transformation and data-driven decision-making as the most important emerging themes in operational excellence. They described how technologies such as artificial intelligence, machine learning, and big data are enabling improvements across business processes. They further argued that these technologies are only effective when combined with strong leadership, a skilled workforce, and a supportive organizational culture.

2.4 Technical Understanding and KPIs

According to Anvari et al. (2010), organizations rely on different performance measurements to evaluate their manufacturing performance relative to competitors and maintain competitiveness. Cristea and Cristea (2021) emphasized that KPIs vary across industries depending on the specific application and process context. They also pointed out the importance of systematically measuring performance indicators and continuously monitoring them over time, highlighting that several KPIs can be used to assess firm performance. Furthermore, they highlighted the value of comparing performance against industry peers to support decision-making and to identify gaps between current and desired performance levels.

OEE is described by Calandreli et al. (2025) as a widely used KPI in the manufacturing industry when an organization aims to assess the efficiency of production equipment or a production process. Furthermore, Anvari et al. (2010) described OEE as a measure of how closely a machine's actual performance aligns with its optimal performance. Ahmad and Dhafr (2002) characterized OEE as an indicator of both equipment reliability and the ability to maintain stable operations over time.

In the literature, OEE is commonly calculated as the product of Availability, Performance, and Quality. However, Jonsson and Lesshammar (1999) note that the exact definitions of these factors vary depending on the application. In addition, they mentioned two different ways of defining the three components, referring to definitions proposed by Nakajima (1988) and De Groote (1995), who conceptualize the components differently.

Jonsson and Lesshammar (1999) described the Availability KPI as a measure of how well actual production time is utilized compared to planned production time, and it can therefore capture the time during which the system is not operating. They continued by describing the Performance KPI as the ratio comparing actual speed to the speed at which the equipment is designed for. They defined the Quality KPI as the proportion of good output relative to total output, focusing on defects generated at the equipment level, excluding downstream quality losses. Jonsson and Lesshammar (1999) further noted that measuring downstream losses as well could provide a more comprehensive view, but it can be complicated to assign responsibility for the defects and their causes.

In addition to the description of the three OEE components, Jonsson and Lesshammar (1999) presented two different ways of defining them, referring to approaches proposed by Nakajima (1988) and De Groote (1995), who conceptualize the components differently.

Nakajima defined Availability as the ratio of loading time (i.e., planned production time) minus downtime to loading time. In contrast, De Groote defined it as planned production time minus unplanned downtime, relative to planned production time.

Regarding Performance, Nakajima defined it as the product of ideal cycle time and output divided by operating time. De Groote, on the other hand, defined Performance as the ratio between actual and planned output, emphasizing alignment with production targets.

For Quality, Nakajima formulated it as the proportion of good output, calculated as the difference between input and defective volume relative to total input. Similarly, De Groote defined Quality as the ratio of accepted production to total actual production, focusing on the share of non-defective output.

Jonsson and Lesshammar (1999) also noted that speed losses within the Performance component are challenging to quantify, as the optimal speed is not always clearly established and can vary depending on factors such as product type, material conditions, and the speed at which the machine is designed for. They further indicated that losses associated with reduced speed result from differences between design speed and actual operating speed. They stressed the importance of using a consistent reference speed for each product when conducting measurements, since variations in this reference make it difficult to compare performance results over time. Similarly, Calandreli et al. (2025) adopts a comparable approach, defining Performance as the relationship between actual and ideal speed, while Quality is expressed as the share of defect-free output relative to the total processed volume.

Calandreli et al. (2025), like other scholars, highlighted that OEE is a useful tool for identifying production losses, thereby supporting the prioritization of improvement initiatives. They outlined six key categories of time losses commonly referenced in OEE literature, originally introduced by Nakajima (1988). These losses, according to Calandreli et al. (2025), can often be categorized as follows: planned downtime (referred to as setup time), unplanned downtime, speed losses, minor stoppages, defects, and yield losses, referring to the time allocated to reworking products that did not meet quality requirements. They further described how each of these loss categories can be linked to the three components of OEE. An overview of these losses and their relationship to OEE is illustrated in Figure 1, based on the framework presented by Calandreli et al. (2025).

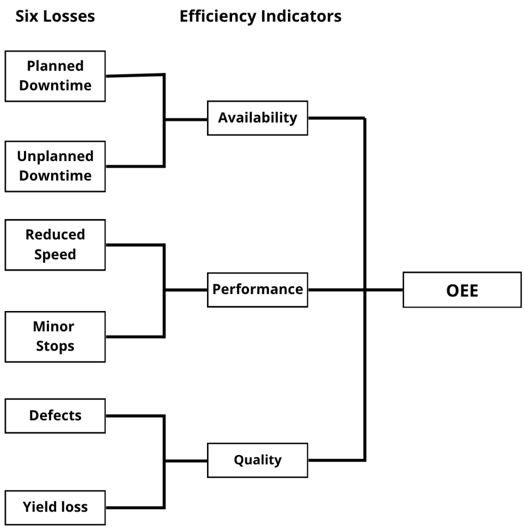


Figure 1: Breakdown of OEE components and associated losses inspired by Calandreli et al. (2025)

Jonsson and Lesshammar (1999) emphasized that meaningful performance improvements depend on well-designed measurement systems and appropriate KPIs. They also pointed out that comparing OEE across different organizations or production sites is problematic, as variations in definitions and operating conditions can result in misleading conclusions. Additionally, they noted that there is an absence of a universally accepted benchmark for OEE, since reported target levels differ across studies.

In contrast, Calandreli et al. (2025) provided a classification of OEE performance, where values below 65% are considered low, while values above 85% are regarded as excellent. Similarly, Ahmad and Dhafr (2002) discussed benchmarks for world-class OEE performance, suggesting that OEE levels above 85% indicate strong performance, with Availability, Performance, and Quality each exceeding 95%. They further highlighted that, because OEE is calculated as a multiplicative measure, achieving near-perfect levels requires all components to perform at very high levels, typically close to 99.5%.

Ullah et al. (2023) highlighted that OEE-based measures of operational performance and effectiveness can lead to different conclusions depending on the perspective applied. They illustrated this by noting that even if an organization's internal OEE does not meet its own targets, it may still perform well relative to industry benchmarks. This suggests that the organization's performance goals may be more ambitious than the standard performance within the industry.

Ahmad and Dhafr (2002) argued that achievable performance levels and targets are strongly influenced by the underlying process technology, which helps explain why relatively low OEE values, often below 50%, are common in process industries. To reach higher performance levels, they emphasized the importance of advanced improvement approaches such as Six Sigma, automation, predictive maintenance, and improved process control.

In addition, Ahmad and Dhafr (2002) suggested that OEE can be improved through practical measures including operator training improvements at bottlenecks, improved production scheduling, product redesign, and more effective operating procedures. They also highlighted that quality improvements can be achieved by increasing right-first-time rates, thereby reducing the need for rework and process adjustments. Furthermore, while some improvement initiatives may require substantial investment, they stressed the importance of using accurate OEE measurements to properly evaluate the impact of such investment-driven improvements.

2.5 Characteristic Dependency Matrix

Erdil and Arani (2018) described Quality Function Deployment (QFD) as a systematic method for converting customer requirements into increasingly detailed technical specifications through a sequence of matrices. The process begins with the House of Quality (HoQ), where customer needs are connected to technical requirements using a relationship matrix. QFD then progresses through several stages, linking design requirements to part characteristics, followed by process and production requirements. Figure 2 presents the third matrix in the sequence, where part requirements are translated into process requirements. In this matrix, the relationships are illustrated through varying symbols, such as circles and triangles, which indicate the strength of the connections.

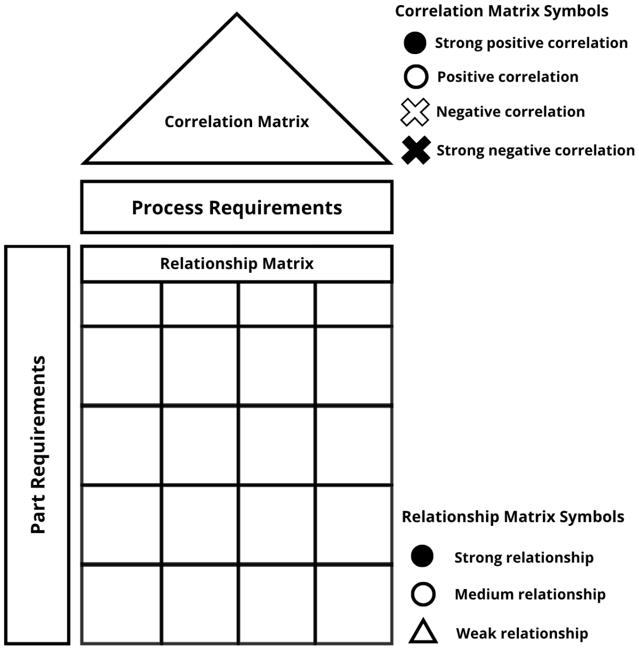


Figure 2: House of Quality (HoQ): Relationship matrix between part requirements and process requirements, inspired by Erdil and Arani (2018)

2.6 Multi-Product Production System

Building on the previously described concept of OEE, Li et al. (2021) argued that the measure was originally developed for a single-product or single-line environment. Consequently, its application becomes more challenging in contexts where multiple products are produced simultaneously on the same equipment.

In a multi-product production system, Li et al. (2021) identified three key characteristics that distinguish it from a single-product environment: organizational complexity, bottleneck variability, and product specification diversity. Particularly relevant is the notion of product specification diversity, where products manufactured on the same line differ in physical attributes such as size, thickness, and processing requirements, leading to different process settings and production speeds across product types. As a consequence of these differences, each product type carries its own theoretical bottleneck speed and its own actual bottleneck speed, meaning that performance varies inherently across products on the same line.

Li et al. (2021) concluded that applying a single shared performance measure across all products in such a system will conceal losses at the individual product level and produce inaccurate assessments of overall system effectiveness. To address this, they proposed that products must be measured separately before any aggregation of system-level performance can be considered meaningful.

While the need to treat products separately in a multi-product environment has been established, the question remains as to what drives the performance differences between products. Trattner et al. (2019) addressed this through a systematic literature review of the relationship between product complexity and operational performance in manufacturing firms, defining product complexity as the number, diversity, and interrelatedness of product variants and components within a production system.

The review found that product complexity has a consistently negative relationship with operational performance, particularly in terms of processing time and productivity. As product complexity increases, processing times lengthen and productivity decreases, meaning that more complex products inherently require more time and resources to produce than simpler ones (Trattner et al., 2019). This relationship holds across a range of industries and production system types, suggesting it reflects a fundamental characteristic of how product complexity interacts with manufacturing processes rather than being context-specific.

Trattner et al. (2019) further argued that industry-specific complexity measures, that is, physical product attributes particular to the production context, provide the most insightful basis for understanding how complexity drives performance differences. Composite or generic measures of complexity were found to produce less conclusive results, whereas detailed, context-specific measures revealed the clearest relationships with operational performance. Additionally, they noted that process industries operating with continuous production systems remain underexplored in the existing literature, representing an area where further research is particularly warranted.

2.7 Product-Mix Based Weighting Approach

This section presents what different sources discuss about weighting in multi-criteria evaluations. Weighting is an important part of any evaluation that includes more than one criterion. According to Greco et al. (2019), a weight shows how important one criterion is in relation to the others. The choice of weights is important because changing them can have a large impact on how units are ranked, and in some cases the results could even be manipulated by choosing different weights. A common approach is to give all criteria equal weights. However, they argued that this is not a good solution because it treats all criteria as equally important, which is rarely the case in reality. They also pointed out that equal weighting is just as subjective as any other weighting choice, even though it might seem more neutral.

2.7.1 Weights Industry Context

Oral et al. (1999) argued that assessing a firm's competitiveness relative to its competitors in a target market can provide valuable insights. They presented various performance measures and illustrated one approach for calculating a firm's total output over a given period. In this approach, total output is expressed as a weighted sum of all products, where each product is multiplied by a conversion coefficient. This coefficient reflects the level of capital resources required to produce one unit of each product. The use of different weights is based on the assumption that products contribute unequally to overall output. Accordingly, the total output is calculated as the sum of the planned quantity of each product j, multiplied by its corresponding α_j, across all products j=1,2,3...

$$Q_F = \max \sum_j \alpha_j x_j \tag{2.1}$$

Furthermore, Oral et al. (1999) explained that operational efficiency at the firm level can be evaluated as the ratio between actual performance and maximum attainable output, seen in Formula 2.2. In their formula, Q_F represents the realized output, while Q_F^* denotes the maximum possible output. Based on this, operational efficiency is expressed as a value between 0 and 1, where values closer to 1 indicate higher efficiency, and values closer to 0 reflect lower performance.

$$h = \frac{Q_F}{Q_F^*} \tag{2.2}$$

In addition, Oral et al. (1999) introduced an external performance measure referred to as Industry Mastery, which evaluated a firm's position within its industry. In this context, Q_R is defined as a relative measure of competitiveness based on external benchmark values from a selected competitor. The index h_{RF}, seen in Formula 2.3, is used to express the ratio between the firm's performance and that of the competitor, thereby indicating its relative standing.

$$h_{FR} = \frac{Q_F}{Q_R} \tag{2.3}$$

Oral et al. (1999) further explained that a value of $h_{RF} > 1$ indicates that the firm performs better than its competitor in the market, whereas a value below 1 suggests weaker performance relative to the competitor.

2.7.2 Production Context Weighting

Puvanasvaran et al. (2013) emphasized that the mix of products being produced influences OEE measurements, which in turn affects the identification of hidden losses, also referred to as hidden costs. They further noted that machine availability is often impacted by frequent changeovers, which arise from variations in the product mix.
To address this, Puvanasvaran et al. (2013) applied a weighted average approach to estimate an aggregated measure of production. This method accounts for differences in product mix by incorporating variations in cycle time and equipment utilization, ensuring that each product contributes in proportion to its actual share of total production time. The weighting factor was determined based on each product's cycle time relative to the overall production time. More specifically, the weight is calculated as the proportion of time spent producing a given product in relation to the total machine run time. Using these weights, an aggregated production measure is then calculated according to Formula 2.4 below.

$$\bar{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \tag{2.4}$$

Puvanasvaran et al. (2013) explained that in this formula, n represents the number of different product types, w_i denotes the constant cycle time associated with a specific product, and $\bar{y}$ corresponds to the aggregated production unit.

Puvanasvaran et al. (2013) further explained that the weighting factor should be applied to both scheduled and actual production time in order to express all data in a common unit. The contributions of all products are then aggregated to reflect their respective impact on overall production. They argued that this approach enables a more representative and accurate calculation of the performance ratio within the OEE framework.

2.8 Case Company-Specific KPI Theory

This section presents company specific theory about the operational setting as well as how the studied KPIs are calculated based on their specific components.

2.8.1 Operational Setting

Each production plant studied in this thesis consists of two main machine types working in sequence. The primary machine forms the first stage of production, and its output is then passed on to the secondary machines, which transform it into finished products. Together, these machines form a complete production line, from raw material to finished product.

This study focuses on only the primary machine, as it accounts for approximately 50 percent of total production volume and therefore has a significant influence on overall plant performance. The primary machine is modular in design, consisting of a series of distinct processing units, where any single module can act as a bottleneck and limit the speed of the entire line.

Given that each plant operates with considerable independence within the company, the efficiency of individual production plants has a direct impact on the overall performance of the organization.

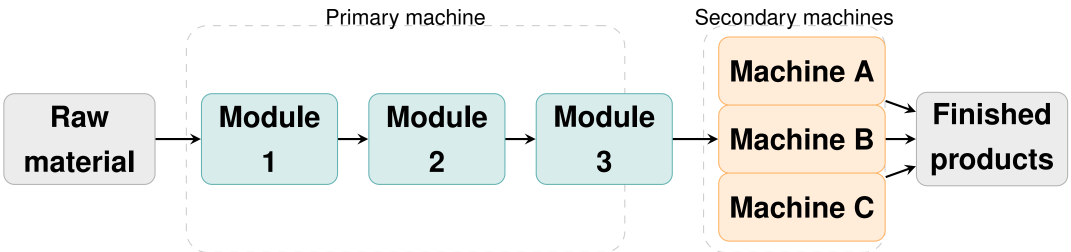


Figure 3: Overview of a production facility

The case company measures plant-level efficiency using OEE. The case company calculates OEE by multiplying three components, time efficiency, speed efficiency, and material efficiency each capturing a different dimension of production loss. Time efficiency and speed efficiency are first measured separately for each machine, and then consolidated into a plant total using weighting factors based on each machine's share of total production volume. Together, the three components give a complete picture of plant performance, which makes the measure useful both for identifying improvement areas and for comparing performance across facilities.

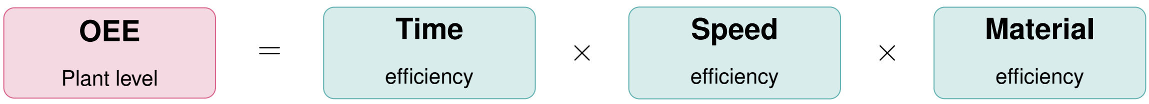


Figure 4: OEE decomposition at plant level

2.8.2 Time Efficiency

The case company measures time efficiency as run time divided by scheduled available time. Scheduled available time is the total time allocated for production and serves as the starting point. From this, three categories of time loss are subtracted to arrive at run time. Setup time covers the time spent on order changeovers and product transitions. Scheduled downtime refers to planned production stops, primarily maintenance activities but also planned interventions such as mechanical adjustments to the production line. Unscheduled downtime captures all unexpected stops where the root cause may not be known in advance. Run time is what remains after these losses.

$$\textbf{Sched avail. time} = \textbf{Setup time} + \textbf{Sched down time} + \textbf{Unsched time} + \textbf{Run time}$$

$$\textbf{Time efficiency} = \frac{\text{Run time}}{\text{Sched available time}}$$

Figure 5: Time efficiency decomposition

Figure 6 illustrates how the time efficiency components contribute to the total scheduled available production time, where the actual run time represents the period during which production takes place.

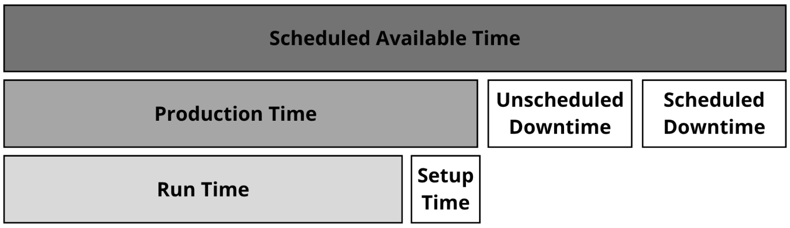


Figure 6: Breakdown of Scheduled Production time

2.8.3 Speed Efficiency

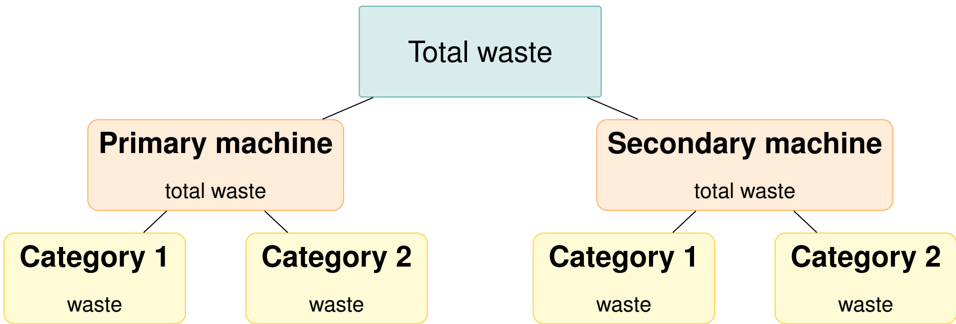
The case company measures the speed efficiency as run speed divided by the machine technical maximum speed. Run speed is the speed at which the machine actually operates, and is calculated as the average actual line speed recorded during run time, without accounting for the fact that different types of products are run at different speeds. The technical maximum speed is the optimal speed for which the machine is designed and therefore determines the maximum capacity of the machine.

$$\textbf{Speed efficiency} = \frac{\text{Run speed}}{\text{Technical max speed}}$$

Figure 7: Speed efficiency decomposition

2.8.4 Material Efficiency

Material efficiency measures the share of raw material that is successfully converted into finished products. It is calculated at plant level by subtracting the total waste ratio from one, where total waste combines waste from both the primary and secondary machines divided by consumed raw materials and sheets, as illustrated in Figure 8. For both machines, waste is divided into two categories. Category 1 waste, which occurs during production as material is adjusted to meet different order specifications. Category 2 waste, which stems from disturbances or interruptions during the production process itself.



$$\text{Material efficiency} = 1 - \frac{\text{Total waste}}{\text{Consumed raw materials and sheets}}$$

Figure 8: Material efficiency decomposition

3

Methodology

In this chapter, the methodology for the research is presented, which includes the research strategy, methods for data collection, methods for data processing, and analysis, followed by a discussion about trustworthiness, reliability, validity, and ethical considerations.

3.1 Research Strategy and Design

The research approach characterizing this study was an abductive approach, which according to Olsson and Sörensen (2021), represents a mix between an inductive and deductive approach. The inductive element supports the mapping of the current state, while the deductive element builds knowledge through existing research. Since the aim of this study was to understand how to evaluate the company's machine performance expressed as KPIs in relation to industry performance, this approach was particularly suitable, as it enabled an understanding of how this could be executed and what characteristics, based on theory, were important to consider. This was achieved through a combination of exploring existing research connected to operations and machine performance, alongside an exploration of the specific machines and how they operate in practice.

The research design of this study was formed around a case study of a company's specific machine type, located across different production plants in Europe. Olsson and Sörensen (2021) described a case study as focused on deeply investigating a specific case, where different data collection methods can be utilized during the investigation. Furthermore, Hammond and Wellington (2021) highlighted that a case study is a flexible research design that can be both structured and unstructured depending on the research need.

A qualitative research method was applied, from an exploratory perspective, with the aim of creating an understanding of the current reality of the machine performance. Hammond and Wellington (2021) highlighted that a qualitative approach is often more inductive in nature with the purpose of exploring and explaining a specific topic. This is consistent with Olsson and Sörensen (2021), who noted that an exploratory approach, which is often qualitative, is aimed at investigating and explaining relationships between different variables. In this study, the qualitative exploratory strategy was therefore applied to create an understanding of the variables affecting machine performance and the relationship between them. Olsson and Sörensen (2021) further described that in a qualitative study, information is often collected from a small number of individuals while investigating several variables, which reflects the approach taken in this study.

In this study, a qualitative method was used to collect both qualitative and quantitative data. Sreejesh et al. (2014) described that quantitative data are generally presented in graphs and charts, which in this study were used to illustrate machine performance. The quantitative data are of a measurable form that would enable a comparison with the internal performance data. While the qualitative data are according to Hammond and Wellington (2021) in a non-numeric form. Regarding the quantitative data collection, some parts of the internal data were already known, while all external data were unknown and had to be collected. The qualitative strategy therefore focused on forming an approach to collect the missing internal and external quantitative data.

3.2 Methods for Data Collection

In this section, the qualitative methods for collecting both qualitative and quantitative data are presented.

3.2.1 Observations

According to Olsson and Sörensen (2021), observations can be of both qualitative and quantitative character. In this study, a direct observation in a qualitative form was conducted, which, according to Olsson and Sörensen (2021), can be described as the observation of situations directly surrounding the researcher. This direct observation was conducted through what is commonly referred to in lean manufacturing as Gemba walks. Márquez Figueroa et al. (2025) described that the concept of Gemba originates from the Toyota Production System and refers to ”the actual place” where value is created in a manufacturing context, this means the shop floor where products are made and processes unfold.

Two plant visits were conducted to plants A and E, where the direct observations through Gemba walks took place. During these visits, the physical production environment, machine layouts, and operational flows were observed, providing an initial solid understanding of the manufacturing context. According to Márquez Figueroa et al. (2025), a Gemba walk is a structured observational practice in which managers or researchers physically walk through the production environment to directly observe workflows, machine operations, and working conditions, with the purpose of identifying improvement opportunities and building an understanding that cannot be obtained from a distance.

3.2.2 Interviews and Questionnaire

The qualitative data collection consisted of different types of interviews as well as a questionnaire.

Internal Exploratory Interviews
Exploratory interviews were conducted in the earlier phase of the research. Sreejesh et al. (2014) highlighted that this is common practice, particularly when discussing problems with experts and analyzing a situation. These interviews were conducted with key personnel from different roles within the organization, including plant managers, technology managers, and production specialists. In total, five exploratory interviews were carried out during a plant visit to plant E. Inspired by Gioia et al. (2012), informants were treated as knowledgeable agents who understand their own organizational reality and can articulate it meaningfully. During the exploratory phase, the interviews focused on exploring the machines, their capabilities, performance, challenges, and factors that could affect machine performance. This approach was particularly suited to the exploratory stages of the research where the goal was to discover rather than confirm existing assumptions.

An advantage of the exploratory approach was that, as stated by Sreejesh et al. (2014), it provided clarity about what type of data needed to be collected for the research to proceed. This exploratory phase was therefore a crucial step in understanding what type of data needed to be collected externally. The data collected during the exploratory phase can be regarded as primary data, as it was gathered directly from the source. According to Sreejesh et al. (2014), primary data are collected directly from respondents. All respondents for the different types of data collection methods are presented in Table 3.1.

Internal Semi-Structured Interviews
After the exploratory phase, several semi-structured interviews were held with internal company personnel, with the purpose of confirming the themes identified during the exploratory interviews and conducting an expert control check of the variables affecting machine performance. Preliminary insights from the initial exploratory phase sug-

gested that products could be categorized based on certain characteristics related to complexity. This was further supported by the literature, where Trattner et al. (2019) indicated that product complexity can be classified into distinct categories based on certain characteristics. These preliminary characteristics were therefore used to structure the semi-structured interviews, where their relevance and validity were further explored and confirmed. This is in line with the description of data triangulation presented by Olsson and Sörensen (2021), in which data from different sources are cross-checked through the use of different data collection methods, to reduce uncertainties and confirm findings.

According to Hammond and Wellington (2021), a semi-structured interview is an interview type where typically the same set of questions is covered, but in an open-ended manner, with the advantage of capturing unexpected information. Hammond and Wellington (2021) highlighted the advantages of this structure, such as avoiding the inflexibility of a fully structured interview while allowing respondents to contribute valuable information that the researcher did not anticipate. These semi-structured interviews were carried out via Microsoft Teams due to differences in location. The data collected through these interviews constitutes primary data, as described by Sreejesh et al. (2014).

External Semi-Structured Interviews
The external data collection was carried out through three steps including an introductory interview, questionnaire, and follow-up interviews. This was also a part of the data triangulation, cross-checking the themes from an external perspective, confirming the themes once again to further reduce uncertainties.

The first step of the external data collection was aimed at confirming the themes discovered during the internal exploratory and semi-structured interviews. This was carried out through introductory interviews with suppliers to establish a common understanding of the characteristics being investigated in relation to machine performance. These interviews were conducted in a semi-structured form, which was chosen due to the advantage described by Sreejesh et al. (2014), who highlighted the flexibility to adapt to different responses. Both suppliers were included in the confirmation of the internal themes. However, for the collection of industry reference data, only one of the suppliers was included, as the other supplier only wanted to contribute to the purpose of understanding the performance characteristics.

The introductory interview with the supplier who contributed industry data was carried out to create a common understanding of the data required to create an external reference level. This interview was structured around a questionnaire created for the purpose of collecting industry data and understanding product characteristics. This can also be referred to as a semi-structured interview as described by Hammond and Wellington (2021), as it followed a structure while allowing the respondent to bring new insights.

External Questionnaire

The second step included the questionnaire which was distributed to the supplier via email. The questionnaire served as a structured and efficient tool for collecting external data, adapting the structure to fit the purpose of the study. It consisted of a fixed number of questions, where some had a fixed range of answers while others were open-ended, providing a high degree of structuring while allowing for contextual elaboration where needed. The structure of the questionnaire can be found in Appendix A. According to Olsson and Sörensen (2021), a high degree of structure and standardization is achieved when a fixed number of alternative answers is provided. However, the questionnaire included a mix of qualitative and quantitative questions, with the purpose of collecting both numeric data and a contextual understanding of that data. One part of the questionnaire was structured around understanding relationships between different themes connected to the KPIs identified during the exploratory phase. To collect this data, the supplier was provided with a relationship matrix inspired by Erdil and Arani (2018) description of the relationship matrix included in the HoQ matrix, to enable a categorization of themes identified in the exploratory phase and their relationships to machine performance.

The questionnaire also included questions about best-in-class and typical performance data for components connected to the studied KPIs for the year 2025 to later compare these with the internal performance data. Distributing the questionnaire allowed the relevant personnel at the supplier to collectively contribute their expertise, consult colleagues, and retrieve data from internal databases. Sreejesh et al. (2014) highlighted that text-based data collection can provide more reflective answers compared to spontaneous verbal responses, as participants have the opportunity to carefully collect and consider the information they provide. This was an important advantage in this context, given the complexity of the data required.

Since the suppliers drew on their own existing records to answer the questionnaire, this element of the data collection can be understood as a source of secondary data, which Sreejesh et al. (2014) described as data that already exists and has been collected by another party. The suppliers were asked to provide data only for production plants in Europe, in line with the delimitation of this study. It should be noted that suppliers, while knowledgeable, represent an external reference point rather than direct industry competitors, making them an indirect source of industry performance data. Collecting data directly from competitors was not a viable option due to the sensitive and confidential nature of production performance data in a competitive market landscape.

External Semi-Structured Interview

In the third and final step, a semi-structured follow-up interview was conducted approximately one week after the questionnaire was distributed, providing the supplier with sufficient time to complete the data collection internally. This interview followed a semi-structured format based on the themes of the questionnaire, while allowing the respondent to contribute any additional and necessary information. These were some of the benefits of this structure described by Hammond and Wellington (2021). This follow-up interview therefore served a dual purpose of verifying the accuracy of the questionnaire responses and allowing the supplier to clarify any uncertainties and contribute any additional information considered relevant. Together, the three steps ensured that the data collected through the questionnaire was both accurate and grounded in a common understanding of the research context, strengthening the overall credibility and reliability of the external data collected.

Internal Questionnaire

An internal questionnaire was created to collect plant-specific data from each of the seven production plants included in the study. This plant-specific data reflected different products in relation to how much of the total production run time is dedicated to each of these different product types. This choice of data to collect from each site was based on insights from the literature review, where Puvanasvaran et al. (2013) stated that differences in product mix were identified as a factor that could influence measured production speed in a multi-product manufacturing system. Unlike the external questionnaire, the internal questionnaire did not require an introductory or follow-up interview, as the data required was more straightforward and did not involve the same complexity of industry-level performance data. The questionnaire was structured around collecting data on production proportions of products and focused specifically on collecting the data needed to enable the integration and comparison of internal performance data with external industry reference data. This plant-specific information was necessary to ensure that the external industry data could be made comparable to the internal performance data in a meaningful way. The internal questionnaire was distributed to and returned from each plant via email, covering plants A through G.

3.2.3 Extraction of internal performance data

The internal performance data for each of the seven machines were collected from an internal Power BI database. This was classified as secondary data, which, according to Sreejesh et al. (2014), refers to data that has already been collected by another party. The data were extracted and structured in an Excel sheet to allow for further processing.

Table 3.1: Interview respondents

Respondent	Role	Interview type	Date
Respondent A	Lean Manager Production	Introductory interview Follow-up questions Questionnaire	9/2-6/5
Respondent B	Plant Manager Production	Questionnaire Follow-up questions	27/4-6/5
Respondent C	Plant Manager Production	Questionnaire Follow-up questions	27/4-6/5
Respondent D	Machine Manager Production	Questionnaire Follow-up questions	27/4-6/5
Respondent Group E: Respondent J, Respondent K Respondent N	Plant Director & CI Manager Production Development Specialist S&OE Regional Manager	questionnaire Follow-up questions	27/4-6/5
Respondent F	Plant Manager Production	Questionnaire Follow-up questions	27/4-6/5
Respondent G	Production Technician	Questionnaire Follow-up questions	27/4-6/5
Respondent H	VP Technical Production	Questionnaire Follow-up interview	27/4-6/5
Respondent I	Technology & Project Excellence Manager	Introductory interview Follow-up interview	2/3-26/3
Respondent J	Plant Director and CI Manager	Introductory interview	3/3

Respondent	Role	Interview type	Date
Respondent K	Production Development Specialist	Introductory interview and follow-up interview	2/3-26/3
Respondent L	VP Technical Production	Introductory interview Follow-up questions	11/3-6/3
Respondent M	Production Manager Production	Introductory interview	2/3-20/3
Respondent N	S&OE Regional Manager Production	Questionnaire (included above)	
Supplier A	Technical Sales Representative Process engineer	Industry Benchmark: Introductory interview Questionnaire Follow-up interview	15/4-23/4
Supplier B	1 Technical Sales Representative 4 Process engineer	Product characteristic Introductory interview	13/4-21/4

3.3 Literature Study

The initial literature search was conducted using a PRISMA-inspired approach. Page et al. (2021) described PRISMA as a standardized framework for conducting and reporting literature searches, built around four defined phases: identification, screening, eligibility, and inclusion. By following this structure, the search process becomes transparent and possible to replicablee.

In the identification phase, searches were performed in Web of Science using the following keywords: ”Overall Equipment Effectiveness”, ”benchmarking”, ”performance improvement”, ”OEE”, ”Industry 4.0”, ”benchmarking framework”, ”key performance indicators”, ”operational performance”, and ”manufacturing”. The initial searches resulted in 1,042 articles. To manage and consolidate the results, the Web of Science Marked List function was used, which allowed all results to be compiled across search strings. After removing duplicates through this function, 222 unique articles remained.

In the screening phase, conference papers and proceedings were first removed, as they are generally considered less rigorous than peer-reviewed journal articles, reducing the pool to 163 articles. Titles and abstracts were then reviewed, and 60 articles were selected for further reading.

In the eligibility phase, the full text of the 60 remaining articles was reviewed. Articles were included if they addressed benchmarking or performance measurement in an industrial context, and excluded if they did not meet this criterion. In the inclusion phase, 8 articles were found to be relevant and included as the theoretical foundation for benchmarking and technical understanding of KPIs. The full selection process is illustrated in Figure 9.

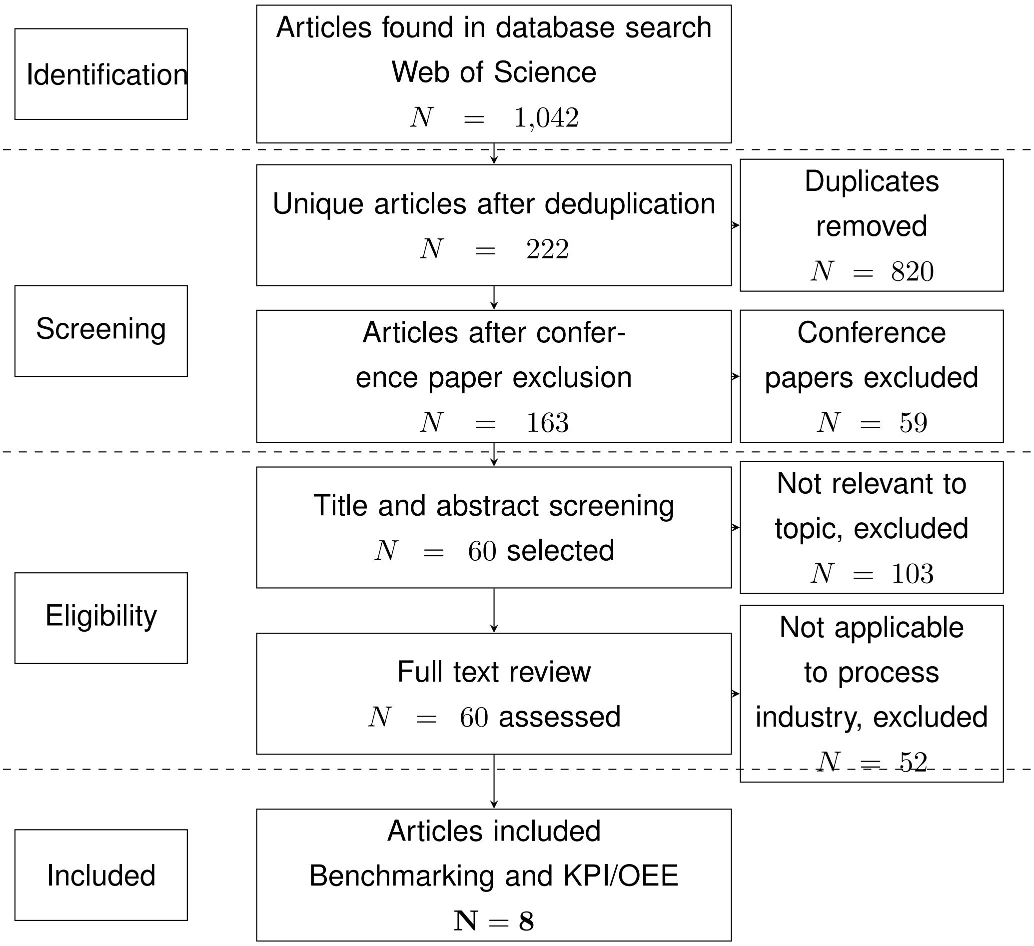


Figure 9: PRISMA-inspired flow diagram for literature search

Despite the initial database search, it became evident that the available literature directly addressing the specific research area of this thesis was limited. Therefore, two complementary search strategies were adopted.

Wohlin (2014) describesd snowballing as particularly valuable when database searches produce limited relevant results, or when the terminology in a field is inconsistent, causing relevant papers to be missed in keyword-based searches. Using the 8 articles identified through the PRISMA-inspired search as a start set, reference lists were systematically examined through backward snowballing in Web of Science and Google Scholar to identify additional sources. Conference papers were excluded. They recommend a stepwise screening process where articles are first assessed based on their title, followed by the publication venue, and then the abstract before a final decision is made based on the full text. Following this process, articles were included if they addressed benchmarking or performance measurement in a relevant industrial context, and excluded if they did not contribute theoretically to the research area. This proved especially useful for identifying foundational and original works within benchmarking and technical understanding of KPIs, where much of the most relevant literature consisted of older works not easily captured through keyword searches.

The second complementary strategy was inspired by the systematic literature review methodology described by Tranfield et al. (2003). They argued that in management and engineering research, inclusion decisions should be based on a study's relevance to the research question and its theoretical contribution rather than strict methodological criteria. Searches were conducted in Web of Science and Google Scholar, and conference papers were similarly excluded. For more specific topics such as multi-product production systems, weighting methods and Porter's Five Forces, searches were therefore conducted with a clear theoretical purpose. Articles were selected based on how well they contributed to the theoretical foundation of each topic, and excluded if they did not sufficiently address the research area in question. Additionally, the literature related to the House of Quality was identified through recommendations. In total, the two complementary strategies resulted in approximately 16 additional articles, which together with the 8 articles from the PRISMA-inspired search form the theoretical foundation of this study.

3.4 Data Processing and Analysis

The exploratory interviews were processed using the Gioia method. According to Gioia et al. (2012), the Gioia method for qualitative research works in three steps. First, the informants' own words and expressions were recorded and preserved as 1st-order concepts, staying close to how the informants themselves describe their reality. These are then grouped into broader 2nd-order themes that capture common patterns across the interviews. Finally, these themes are condensed into aggregate dimensions that together provide a structured overall understanding. Throughout this process, the researcher avoids imposing existing theories or assumptions on the informants, and instead lets the key concepts emerge naturally from the conversations. They described that this approach is particularly suited to the exploratory stages of research where the goal is discovery rather than confirmation of existing assumptions.

In this study, the transcribed interview material was read through and the respondents' own words were preserved as they were expressed. These were then grouped into broader themes that captured common patterns across the interviews, providing a strong basis for the subsequent analysis. The semi-structured interviews, conducted with both internal personnel and external respondents, were summarized and organized according to the same themes identified during the exploratory phase.

The data processing of the themes identified through the interviews was further carried out through a systematic mapping of characteristics and categories Grounded Theory method described by Olsson and Sörensen (2021), which implies that data are systematically grouped and mapped based on categories and subcategories. The qualitative data were analyzed based on the different variables identified within each theme, including their subcategories and other variables that had a relationship with the different variables. This enabled an understanding of what can cause differences in performance. This approach to mapping was further inspired by Slack and Lewis (2015), who explained the decomposition of a ratio and how it drives areas connected to operations strategy, providing a structured framework for understanding how individual variables relate to overall performance outcomes.

The quantitative data were analyzed by comparing the internal performance data with the integrated external industry reference data to identify gaps in performance. The industry data provided by the supplier was processed to be integrated and compared with internal performance data by using formulas inspired by the ones presented by Puvanasvaran et al. (2013) and Oral et al. (1999) in the literature, together with the theory about the company KPIs found in section 2.8.

When evaluating performance metrics against reference values, bullet charts were used to visually illustrate these gaps. Hardin et al. (2012) recommended bullet charts for this purpose, as they allow a primary measure to be visually compared against one or more reference measures, with color used to indicate levels of achievement relative to the reference values.

Figure 10 illustrates the research process as a flow chart, including each step taken to collect data.

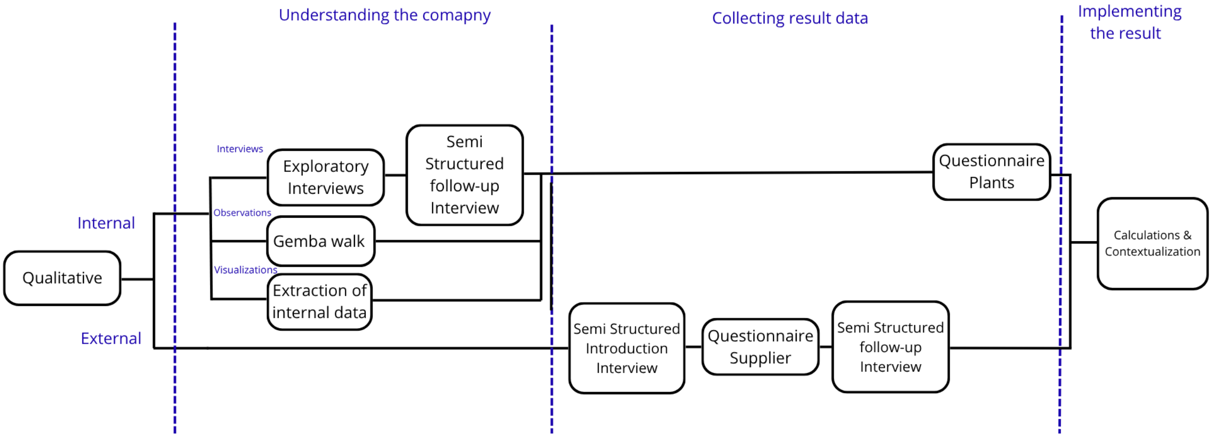


Figure 10: Research methodology overview

3.5 Trustworthiness

To strengthen the trustworthiness of the study, data triangulation was applied throughout the data collection process. According to Olsson and Sörensen (2021), data triangulation refers to data being collected in different ways. Hammond and Wellington (2021) emphasized that the benefit of mixed methods is to provide confirming or complementary sources of data.

In this study, data triangulation was achieved through collecting data from a variety of respondents, from key internal personnel to external suppliers with industry knowledge. The data triangulation ensured that the findings of this study were confirmed and complemented across multiple respondents, which strengthened the trustworthiness of this study. This is also what Saunders et al. (2009) described to be the benefit of triangulation, where different independent sources of data can strengthen the trustworthiness of the findings.

3.6 Generalization

According to Olsson and Sörensen (2021), generalizing the results of a study concerns whether the findings can be applied in a context other than the one studied, and they note that this can limit the conclusions that can be drawn. For this study, the results are case-specific, as the themes, variables, and performance values identified are tied to a specific machine type, a specific company, and a specific time period, which limits the direct generalizability of the findings to this particular case. However, a distinction can be made between the generalizability of the results and the transferability of the method. The approach applied in this study, combining internal performance data with externally collected industry reference data through expert interviews and questionnaires structured around identified themes, could be transferred to other machine types, companies, or industries. In such cases, different characteristics and themes would likely emerge, reflecting the specific context of that study, but the overall methodological approach would remain applicable. The study was further limited by data availability, as the external benchmarking data reflects industry estimations rather than exact operational figures from individual plants. Both internal and external data were treated confidentially, which limits the level of detail that can be reported in this thesis.

3.7 Reliability

To strengthen the reliability of the study, contextual detail was prioritized throughout the data collection process, which, according to Hammond and Wellington (2021), is particularly important in small case studies where the purpose is to provide reliability rather than generalizability. The exploratory interviews were processed by identifying common themes across respondents, ensuring that the analysis was based on patterns in the data rather than the subjective interpretation of the authors. To further strengthen reliability, the themes identified during the exploratory phase were confirmed through several internal and external interviews, providing an expert control check of the variables identified. Two plant visits were conducted to deepen the understanding of the production process. The study was kept at a holistic level to enable the examination of several machines across multiple plants, which defined the scope of the data collection. The data collected reflected the performance during a specific time period, and the data collection was therefore structured to capture the most recent available data for the year of 2025.

The reliability of the literature review was also considered. The PRISMA-inspired search was conducted in Web of Science with clearly defined keywords, making the search process transparent and replicable (Page et al., 2021). For the complementary literature search, snowballing (Wohlin, 2014) and systematic literature review (Tranfield et al., 2003) were used as guiding frameworks, both of which provide clear structures for how articles are identified and selected. It was acknowledged that the available literature directly addressing the specific research area was limited, and older foundational works were therefore included to ensure a sufficient theoretical basis.

3.8 Validity

Validity concerns the extent to which a research method measures what it intends to measure (Olsson & Sörensen, 2021), or in other words, the degree to which the research accurately reflects the reality it seeks to capture (Hammond & Wellington, 2021). In this study, validity was strengthened by conducting the exploratory interviews with key personnel directly involved in the operation and management of the machines, including plant managers and technology managers. This ensured that the variables identified during the exploratory phase were grounded in the actual operational reality of the studied machines, rather than being affected by the perceptions of the researchers.

The themes and variables identified during the exploratory phase were validated through several internal and external semi-structured confirmation interviews. This expert control check ensured that the variables being measured were relevant and accurately reflected the factors affecting machine performance, strengthening the content validity of the study.

The product mix proportional weights were collected through the internal questionnaire, where each plant manager made a subjective assessment of their production time distribution based on historical run time data from 2025. Edwards and Barron (1994), who developed a structured approach to multi-criteria decision making, argued that the most knowledgeable expert should make judgments for each dimension, which is why plant managers were asked to provide these weights, as they have direct knowledge of their own production conditions. Greco et al. (2019) noted that expert-based approaches produce weights that are relevant in context, while data-driven approaches reduce subjectivity by basing weights on actual data. In this study, both are present, as plant managers used their own expertise and judgment to classify and distribute their production time across the complexity scenarios, while historical run time data from 2025 served as the objective foundation for that assessment.

The same logic applies to the external supplier questionnaire, which was answered by sales managers and process engineers, people with hands-on knowledge of industry conditions. The questionnaire mixed quantitative questions based on historical performance data with qualitative questions that required expert judgment, reflecting the same combination of data-driven and subjective judgment elements described by Greco et al. (2019). This means the data collected are both grounded in real numbers and interpreted by people who actually understand the context. Edwards and Barron (1994) also pointed out that working with a small group of highly qualified experts is better than involving many, since larger groups tend to introduce inconsistency. The choice to keep the respondent group small and knowledgeable therefore strengthens rather than weakens the validity of the data collected.

Validity was also supported through the introductory and follow-up interviews conducted with the supplier, which ensured that a common understanding of the questionnaire questions and responses was established, reducing the risk of misinterpretation of the data collected.

Regarding the literature review, the validity was supported by basing the theoretical framework on peer-reviewed journal articles and well-established foundational works within the relevant fields. By using structured search methods including PRISMA, snowballing, and a systematic literature review, the risk of overlooking relevant literature was reduced. However, as discussed, there is limited literature available that directly addresses external benchmarking through machine suppliers. This means that the theoretical framework partly relies on more general sources that may not fully capture the specific context of this study.

3.9 Ethical considerations

Each respondent was informed of the purpose of the study and how the information they provided would contribute to it. All respondents were anonymized, and the data collected was treated with confidentiality. During the interviews, respondents were asked whether it was acceptable to transcribe or record the interview, and if anyone declined, it was handled according to their wishes.

The data presented in this report primarily reflect proportions rather than exact figures, as exact figures are considered sensitive. Where numerical values have been included, they do not represent the actual figures, instead they have been modified to provide the reader with an understanding of the context. Due to ethical considerations, data collected in measurable and computational form was not presented in the graphs in this thesis, instead, the primary relationships between the data were shown, without revealing any specific values. If a respondent requested that specific data not be included, that data was excluded from the thesis accordingly. An NDA was signed and followed to not disclose any confidential data.

3.10 Usage of AI

Generative AI tools were used only to refine the language of this study, including grammar and flow of the text. The content, analysis, and conclusions presented are entirely the author's own work.

4

Results

This chapter presents both qualitative and quantitative results through text and supporting graphs.

4.1 Results from Interviews

This section presents the qualitative results of the interviews.

4.1.1 Speed Efficiency and Product Characteristics

The factors affecting speed efficiency were investigated, with the main finding being that it depends on the type of product variation produced.

Factors Affecting Machine Run Speed
According to Respondent M, speed efficiency depends on the type of product being produced. The respondent further noted that order length also affects machine speed, in other words, the length of the time spent running a specific customer order. Respondent I further added that multiple variables affect speed efficiency, including product design, as well as technical limitations, and insufficient operator knowledge regarding when to increase run speed.

Respondent I noted that both the size of the machine and the age of the equipment affect run speed. In this context, Machine Supplier A explained that the maximum speed at which the machine can operate is largely determined by its design specifications, where newer machines are generally capable of achieving higher speeds than older ones.

Additionally, Supplier B noted that machine run speed is strongly influenced by each plant's production mix, that is, how production time is distributed across different product types during manufacturing. Respondent I pointed out that benchmarking speed in an industry context requires consideration of several additional factors, including product material composition, technical conditions, order size, product mix, and the quality of the raw material.

During a visit to plant A, Respondent A stated that a product is composed of different characteristics that affect production complexity. The respondent also mentioned that each product has a structure consisting of different characteristics such as thickness, weight, and material composition. During a plant visit to Production Plant E, Respondent J pointed out that their site runs a high variety of products, with approximately 70 different combinations. Furthermore, the respondent pointed out that product thickness has the greatest impact on machine run speed, noting that thickness is measured in terms of grammage (kg). This was consistent with the characteristics identified by Respondents I, A, and J, with Respondent L additionally highlighting product width as a relevant factor.

Respondents K and L also highlighted that each machine is designed for a specific maximum speed. Respondent K further elaborated that speed efficiency is less affected by product width than by output productivity, explaining that wider products result in greater output per meter of machine run. Furthermore, the respondent elaborated that speed efficiency reflects how effectively the machine operates, assessed by comparing the achieved run speed against the reference speed. Respondent K further identified product structure and material composition as key determinants of run speed, noting that greater thickness reduces the achievable speed, while heavier products generally result in lower run speeds and lighter products result in higher run speeds. The respondent therefore highlighted that an average grammage value can serve as a representative indicator when benchmarking. The respondent further noted that programme setup count and order length also influence run speed, with technology identified as an additional contributing factor. The respondent also noted that although internal speed data is relatively straightforward to obtain, acquiring comparable data from competitors is significantly more complex, given the lack of a standardized approach to how product characteristics affect run speed.

Respondent I stated that defining simplified product scenario specifications could facilitate raw data collection for the benchmarking model. The respondent further highlighted the importance of establishing interval ranges for each product characteristic, such as a grammage interval or a thickness interval, given that product characteristics are not harmonized across production plants. The respondent additionally noted that the type of raw material used and other internal factors may also constrain machine performance. The qualitative findings that laid the base for the benchmark approach regarding the speed efficiency are illustrated in Figure 11 below, showing that machine run speed, which is a part of the speed efficiency calculation, is dependent on product characteristics and complexity. Product characteristics and complexity are captured through four key characteristics: grammage, product width, product thickness, and product structure, each affecting run speed differently. The reference speed, also called the technical maximum speed, is determined by machine design and capabilities. The figure summarizes these findings as factors affecting the components of the speed efficiency Equation.

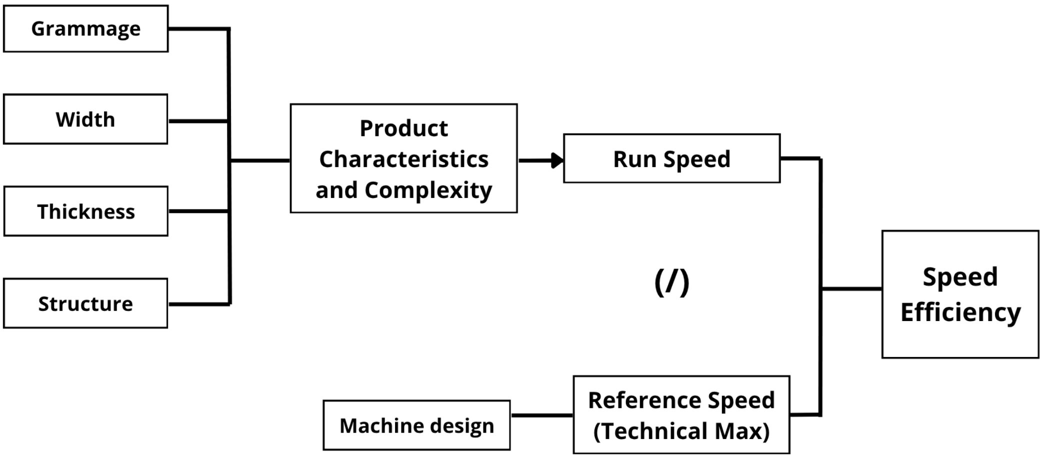


Figure 11: Breakdown of factors affecting machine speed efficiency

Benchmarking Industry Speed

Supplier A highlighted cultural differences between countries, noting that these might influence how operators run the machine. The supplier explained that in certain contexts, operators tend to gravitate towards over time a speed that is more comfortable for sustained operation rather than pushing for the maximum achievable speed. The supplier stated that this tendency is less observable in other parts of Europe, where managers are generally more attentive to understanding the reasons behind speed reductions. The supplier further emphasized that middle management plays a critical role in achieving higher machine speeds, as it is at this level that operational issues are addressed. Furthermore, the supplier stated that numerous factors influence the speed at which a machine operates, highlighting operator knowledge and training as particularly significant in addressing operational issues directly rather than adopting temporary workarounds.

Regarding industry run speed, Supplier A explained that the machines they supply have a maximum rated speed of 1270 (m/min), with some customers achieving this speed, though not consistently. Supplier A stated that 1220 (m/min) represents best-in-class performance, noting that even 950 (m/min) is a speed that very few plants are capable of sustaining continuously. However, 1220 (m/min) achievable provided that the plant operates machines that are not older than five years. The supplier continued by explaining that while some customers are capable of achieving 950 (m/min) when running an order across one or several shifts, this speed cannot be sustained on a daily basis, given that their product mix does not consistently permit such speeds. In addition, the supplier highlighted that typical industry run speeds are often in the range of 680 (m/min), while speeds as low as 460 (m/min) have also been observed, as well as speeds exceeding 800 (m/min).

4.1.2 Time Efficiency

Respondent I stated that

> ”time efficiency has a greater impact on productivity than speed”.

The respondent further elaborated that ”time efficiency can be affected by how you are running the machine, if you have jams, stoppages, or breakages, and unplanned stops could also be due to maintenance or cleaning”. Respondent M stated that ”one thing to improve time efficiency is to regularly carry out scheduled technical maintenance”. Respondent J, on the other hand, noted that the frequency of product structure changes also affects time efficiency. However, Supplier B argued that time efficiency is not determined by any specific product specification and highlighted that

> ”When we speak about efficiency, we mean electro-mechanical availability of their machines during the production time”.

4.1.3 Scheduled Downtime

According to Supplier A, scheduled downtime should be prioritized, though the optimal level depends on the perspective of the respondent. The supplier further explained that machine suppliers advocate for maximizing scheduled downtime to ensure adequate maintenance, whereas manufacturing companies generally seek to minimize it in order to maximize productive machine time. According to the supplier, there is a trade-off between maintenance cost and productivity, which in turn affects the operational lifespan of the machine, where they recommend a specific number of hours of maintenance per year. The supplier further stated that for a machine operating on a three-shift basis, a scheduled downtime of one shift is recommended, which represents a certain percentage of the total scheduled available time. However, the supplier pointed out that this does not reflect common practice among their customers, as machines typically stand still during only half a shift per week, which was identified as insufficient for adequate maintenance. The supplier also observed a trend over the past two to three years towards higher levels of scheduled downtime within the industry, attributing this shift to a growing recognition among manufacturing companies that insufficient maintenance represents a significant operational risk. The supplier stated that increasing scheduled downtime could reduce unscheduled downtime, as it provides an opportunity to identify and address potential issues before they escalate into unscheduled stoppages, thereby enabling problems to be resolved proactively during planned maintenance windows rather than reactively during unplanned downtime.

4.1.4 Unscheduled Downtime

Supplier A clarified that they are able to provide industry benchmark figures for technical unavailability rather than unscheduled downtime. The supplier defined technical unavailability as

> ”planned production time which cannot be utilized due to technical issues or disturbances”.

The supplier continued by explaining that certain factors are excluded from this definition, including, planned maintenance, non-scheduled production time and production or process-related issues. In contrast, the unplanned downtime rate considered relevant for benchmarking purposes was described by Respondents A and I. They explained that unscheduled downtime could be caused by maintenance, cleaning, technical issues, or lack of operator competence. Supplier A stated that the absolute best technical availability in the industry is an extremely high percentage, while the industry standard is slightly lower. The supplier explained that the technical availability implies the machine should be available for a specified percentage of the scheduled production time, during which it is capable of continuous operation in the absence of technical failures, thereby eliminating the need for unplanned downtime. According to Respondent K, unscheduled downtime can be reduced through improved operator knowledge of how to respond to different operational situations, combined with effective communication and collaboration across the team.

In Figure 12 the factors influencing scheduled and unscheduled downtime identified through interviews were mapped out in relation to the time efficiency Equation. This figure can be considered a summary of the findings related to time efficiency.

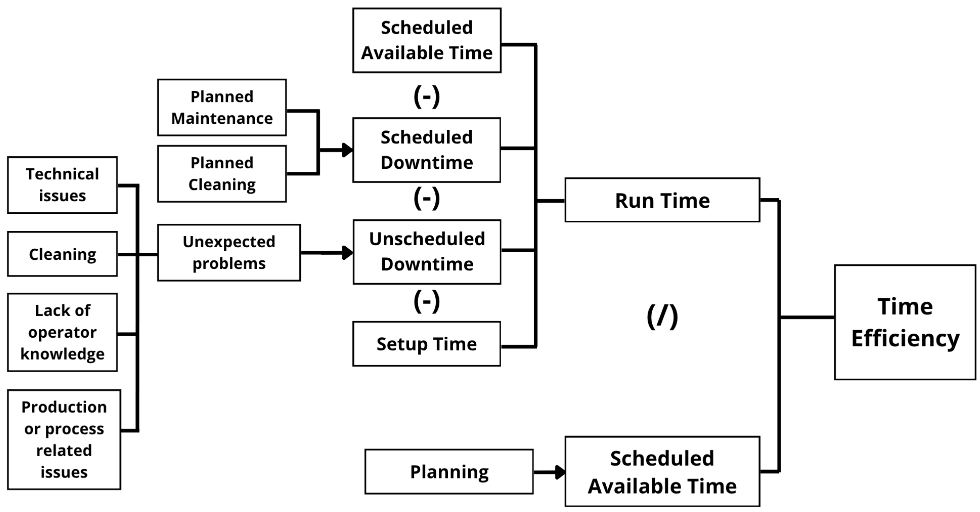


Figure 12: Breakdown of factors affecting machine time efficiency

In Figure13, the factors affecting technical availability are illustrated below, where technical unavailability is calculated by subtracting technical availability from 1. Technical availability is not fully a part of the Time Efficiency KPI, but can be used to make a judgment of the machine's technical capability.

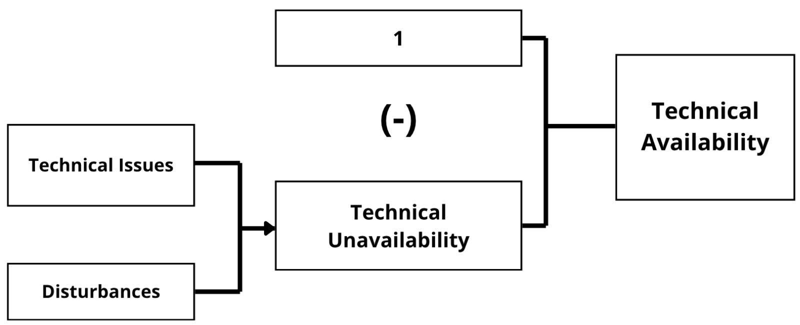


Figure 13: Breakdown of machine technical availability and technical unavailability

4.1.5 Waste

This section presents the understanding of the two waste categories. Supplier A noted that waste in production is not influenced by product type in the same way as run speed, but is instead determined by other factors such as production planning.

Category 1 waste
According to Supplier A, the best-in-class level of Category 1 waste is a quite low percentage, and the supplier explained that this type of waste is heavily dependent on production planning. Respondent M further stated that production planning is largely determined by the type of orders placed by the customers.

Category 2 waste
Supplier A described that waste is measured by their customers in different ways, where some of them do not measure their waste levels, while others only measure the waste generated at the output end of the machine. It was further noted that poor quality may also be detected at later stages of the production process, even if it originates from the primary machine. This may result in waste being incorrectly recorded, leading to misleading statistics. The industry best-in-class Category 2 waste level was a certain percentage of the input material. The supplier described that this level is calculated based on rates from plants recording waste detected at later stages of the production process, caused by the primary machine, since this provides the whole picture. The supplier stated that this is called a registration of base-machine waste even if it is identified later in the process. However, it was found that the case company only measures the waste identified at the end of the machine.

4.1.6 Targets and Motivation

Respondent M described that they visualize their targets on a dashboard next to the machine available for the operators to see. On the dashboard, they display the previous week's and current month's results for speed, time, and material efficiency. In addition, they display last year's target, a value against which they aim to improve their performance. Respondent I stated that the performance targets are established each year based on past year's targets with a percentage increase. In addition, Respondent I pointed out the importance of pushing performance levels by stating that

> "orders and products might not be the same as five years ago and therefore the production plants can be biased regarding what they actually can perform".

The respondent further highlighted the need for an external perspective to motivate continuous performance improvement. In addition, Respondent I noted that operator knowledge and confidence with which operators run the machine represent one of the most significant constraints on performance improvement. The respondent further emphasized that the shift leader or production manager plays a critical role in supporting operators to develop the confidence required to increase run speed.

During a Gemba walk conducted at one of the production plants, an operator was asked why the machine was not being run at a higher speed. The operator explained that increasing run speed results in higher waste levels, noting that product complexity prevented further speed increase. Respondent I stated that improvements in speed and time efficiency can be achieved through measures that do not necessarily require substantial capital investment. The respondent further noted that knowledge sharing between production managers was identified as an additional factor, which the respondent believed could positively influence operator motivation.

4.1.7 Work-In-Process Storage

According to Respondent I, variations in the logistical setup surrounding the machine may also influence the achievable run speed. The primary machine may be required to switch between product structures in order to ensure the continuity of operations on the secondary machine. The respondent explained that where a large work-in-process area exists within the production facility, the primary machine may dictate the overall speed of the production flow. Warehouse capacity and order sizes are frequently identified as limiting factors, and since these are largely beyond direct operational control, focus should be directed towards optimizing the factors that are within the site's capacity to improve. According to Respondent J, if certain components of the internal logistics infrastructure are insufficient, this might result in stock accumulation, ultimately requiring the machine to be slowed down or stopped to prevent further stock overflow.

4.1.8 Machine Automation and Design

Supplier A mentioned that their machines are designed to produce the relevant product range at the highest achievable run speed for each respective product type in that range. This is called the sweet spot, which reflects the machine's design, optimized to efficiently produce the most common product on the market. Furthermore, Supplier A stated that they have seen a shift in the industry due to higher raw material prices, where manufacturing companies experiment with different types of lightweight products, which are not optimal for machine runnability. In addition, the supplier stated that older machines are designed for heavier products, which has resulted in increased production complexity for the older machines. Supplier A also pointed out that product weight varies depending on the intended application and may therefore differ between regions, which in turn influences the level of production complexity.

According to Respondent K, argued that best-in-class is determined by the capability to identify and resolve unforeseen operational issues and the quality of the system in place to detect hidden problems. Mechanical, electrical, and raw material-related failures are inevitable in any production environment, and plants equipped with advanced monitoring systems, such as cameras and sensors, combined with highly skilled operators, are better positioned to identify root causes rapidly, thereby achieving superior efficiency in terms of both speed and time.

Supplier A stated that the latest machines available on the market require fewer operators due to their high degree of automation. The supplier further explained that communication tools facilitating interaction between operators, as well as camera monitoring systems, also play a significant role in operational efficiency. Furthermore, the supplier noted that running highly complex products typically requires the deployment of an additional operator. Figure 14 below summarizes additional factors affecting machine performance identified through interviews.

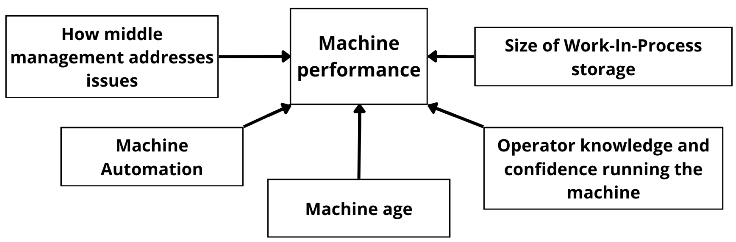


Figure 14: Breakdown of factors affecting machine performance

4.2 Qualitative and Quantitative Results from the Questionnaires

This section presents the qualitative and quantitative results from the internal and external questionnaires in the form of a relationship matrix, scenario interval definitions, and product-mix weights per machine.

4.2.1 Relationship Matrix

During the data collection process, Supplier B was willing to share data, but only under strict conditions to protect the interests of their existing customers. As a result, the data provided by Supplier B was not used, however, it was used to confirm the product characteristics. Supplier A therefore served as the main source of industry benchmark data throughout this study.

The purpose of the matrix presented in Table 4.1 below was to identify which product characteristics have the greatest influence on the studied KPIs, speed efficiency, time efficiency and material efficiency. This would enable an understanding of where an interval was needed for defining and categorizing products into the different scenarios simple, medium, and complex. For instance, certain grammage intervals may allow for higher machine run speeds than others, directly affecting speed efficiency.

Through an introductory interview with Supplier B, it was established that time efficiency was not influenced by the type of product being produced. Consequently, time efficiency was excluded from the relationship matrix. The remaining KPIs were decomposed into their components: run speed as a component of speed efficiency, and Category 1 waste and Category 2 waste as components of material efficiency. This decomposition allowed for a comprehensive understanding of how specific product characteristics influence each KPI component.

Supplier A was asked to assess how each product characteristic, including grammage, width, thickness, and structure, as shown in Table 4.1, affects these KPI components. While doing this, the supplier was asked to take into consideration both the strength and the direction of each effect. The strength of each relationship was expressed on a scale from 1 to 5, where 1 indicates a very weak effect, 2 a weak effect, 3 a moderate effect, 4 a strong effect, and 5 a very strong effect. The direction of each relationship is indicated by an arrow, where $\downarrow$ denotes that an increase in the product characteristic reduces the KPI component, and $\uparrow$ denotes that an increase in the product characteristic increases the KPI component. The symbol $\sim$ indicates a non-linear relationship where the direction depends on additional factors. The run speed row, marked with an asterisk (*), reflects how run speed itself affects the waste categories and therefore contains no entry under the run speed column. Supplier A's responses for each characteristic are presented below.

Grammage:
Supplier A explained that there is a ”sweet spot” for grammage, estimated to be around X-Y g/m2, where machine performance is optimal. Products with a grammage below or above this range negatively influence performance, which is why run speed is indicated by both$\downarrow$ and $\uparrow$ in Table 4.1. Products produced within the ”sweet spot” range consistently yield the best operational conditions. Regarding waste, Supplier A noted that waste is measured in tons rather than in meters of production. Grammage was assessed to have a weak influence on the waste categories overall. For Category 2 waste, the effect is weak but the direction is positive, indicated by $\uparrow$, while Category 1 waste exhibits a non-linear relationship with grammage, where the direction of the effect is dependent on additional factors and therefore indicated by $\sim$.

Width:
Product width does not significantly affect machine performance in terms of run speed, as seen in Table 4.1. Supplier A explained that the machine can run slightly faster and with marginally less waste during the startup phase when running a narrower width, but once the machine is running steadily, the width has no notable impact on performance. Regarding Category 1 waste, it is normally minimized through production planning by taking product widths into consideration in advance. Consequently, Category 1 waste is not significantly affected by changes in product width, as this factor is already accounted for in the planning process.

Thickness:
According to Supplier A, thickness does not affect Category 1 waste, as this is more dependent on production planning. However, an increase in thickness would increase the grammage counted as waste, resulting in a moderate effect on Category 2 waste. Supplier A further noted that thickness is largely influenced by the structure of the product, and therefore both structure and thickness have a strong effect on run speed, with a direction explained by the ”sweet spot” principle. The effect of structure on Category 1 waste is weak, while its effect on Category 2 waste is moderate, as seen in Table 4.1.

Structure:
Supplier A described that different structure types present different operational challenges, and no linear relationship can be used as a reference in this case, yet structure still has a strong effect on machine run speed. The supplier provided an example and explained that one structure type that is thicker and another that is thinner may both run faster than a third structure type due to other machine-related challenges, which explains the non-linear direction indicated by $\sim$ in Table 4.1.

Table 4.1: Relationship matrix between product characteristics and KPI components

	Run speed		Category 1 waste		Category 2 waste	
Product characteristic	**Strength**	**Direction**	**Strength**	**Direction**	**Strength**	**Direction**
Grammage	3	↓↑	1	∼	2	↑
Width	1	∼	1	∼	1	↑
Thickness	4	↓↑	1	∼	3	↑
Structure	4	↓↑	1	∼	3	↓↑
Run speed *	—	—	1	∼	4	↑

4.2.2 Scenario Interval Definition

Based on this relationship matrix, Supplier A was asked to classify product types into the complexity scenarios, simple, medium, and complex, based on the identified characteristics, with runnability and run speed in mind. It was identified that structure, grammage, and thickness all had a significant effect on run speed, while product width did not impact performance in the same way. Therefore, the same interval was applied across all product scenarios. It was further identified that structure and thickness are interdependent, and that thickness alone does not determine whether a product is easy to run. Rather, it is the combination of characteristics that determines the complexity of the product.

The scenarios were constructed to include a range of structure types that vary in runnability, from structures that are easy to run to those that require a reduction in machine speed, as well as defined grammage intervals. Product width was kept constant across all scenarios following a discussion with Supplier A, while thickness was determined by the type of structure. The definitions of the product scenarios based on runnability were as follows:

Simple:
A straightforward product scenario that runs close to maximum machine speed, requiring little or no speed reduction from rated speed.

Medium:
A product scenario that requires some operational adjustment, including a moderate speed reduction, and is still manageable with normal production flow, though at a lower speed.

Complex:
A product scenario that requires significant speed reduction due to the high complexity of the product, leading to challenging production conditions.

The intervals representing the simple, medium, and complex scenarios were defined by Supplier A, who provided concrete value ranges specifying what falls within each complexity level. This meant that the categorization was grounded in measurable, product-specific parameters, making it applicable across the production facilities included in the study. The characteristic intervals defining each scenario are represented in Table 4.2, where each scenario represents a combination of the respective characteristic categories.

Table 4.2: Product complexity scenarios

Characteristic	Simple	Medium	Complex
Grammage	($<$ X gsm)	(X–Y gsm)	($>$ Y gsm)
width	A – B	A – B	A – B
Thickness	2	2	1
Structure	X and XY	O and PO	T and J
Overall description	Can run close to maximum speed and requires minimal operational adjustment.	Manageable within normal production flow and requires moderate speed reduction.	Challenging combination of characteristics which requires significant speed reduction.

4.2.3 Machine Product-Mix Weights

Based on the material provided by Supplier A, the complexity scenarios could only be defined in relation to speed efficiency and not for the remaining KPIs, time efficiency and material efficiency. Therefore, the focus was on collecting data for the weights associated with each run speed scenario based on the product specification scenarios. For the three scenarios, each production plant was asked to provide ratios for their primary machine based on historical run time data from 2025. This data was used to reflect how much of their total run time in production each product scenario, simple, medium, and complex, represented. These ratios represent the product-mix weights, which were used to calculate a weighted industry run speed based on each plant's product mix, simulating how the machines would perform given their specific product mix and the run speeds for each product scenario provided by Supplier A. The distribution of simple, medium, and complex products for each plant is shown in Figure 15 and can be illustrated using the Equations below, where T_i denotes the run time dedicated to scenario $s = simple$, $m = medium$, and $c = complex$.

$$w_s = \frac{T_s}{T_s + T_m + T_c} \tag{4.1}$$

$$w_m = \frac{T_m}{T_s + T_m + T_c} \tag{4.2}$$

$$w_c = \frac{T_C}{T_s + T_m + T_c} \tag{4.3}$$

$$w_s + w_m + w_c = 1 \tag{4.4}$$

The resulting product mix distribution across all seven plants is shown in Figure 15.

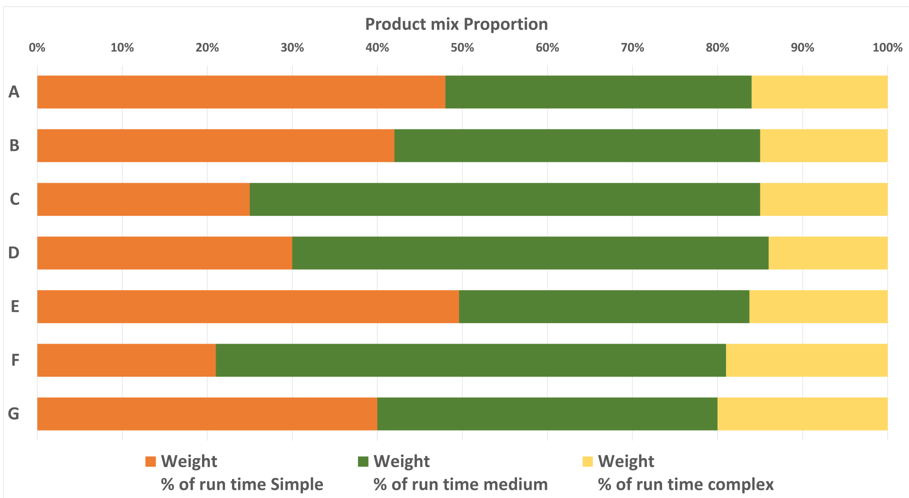


Figure 15: Product mix proportions by plant/machine and complexity scenario (%)

4.3 Proposed Benchmark Metrics

This section presents a simulation of how each plant's actual values relate to industry reference levels, based on calculations using both internal data and external industry data. The aim is to enable a fair comparison between each plant's current performance and the industry reference level, thereby identifying performance gaps. All values are based on data from 2025. Simulation in this thesis is referred to as a way to investigate and test how the company's machines would perform using the industry's reference values.

4.3.1 Speed Efficiency

To benchmark machine run speed externally and compare it against internal performance, the process was carried out in several steps, which are illustrated in Figure 16.

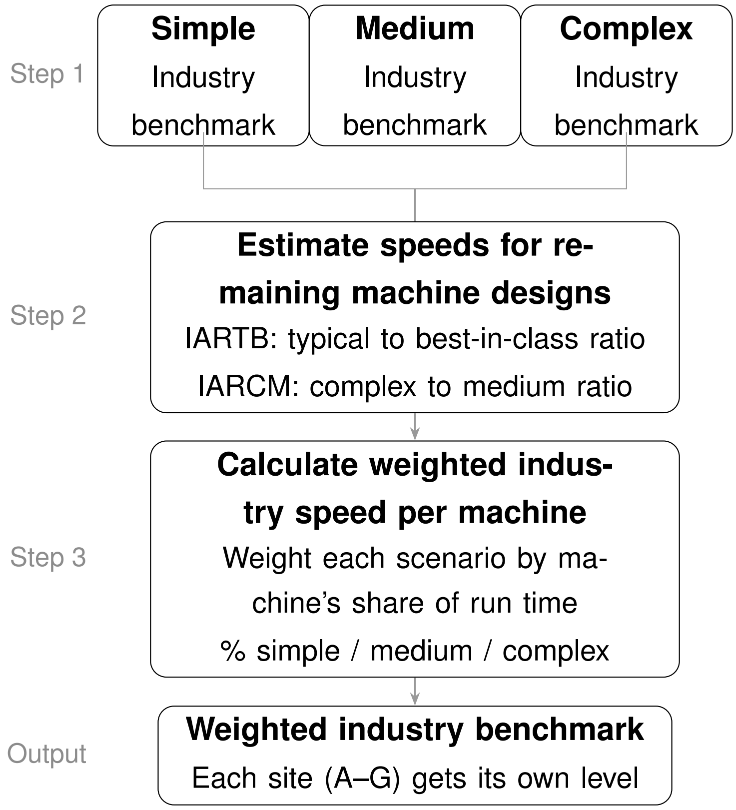


Figure 16: Methodology for constructing the weighted industry benchmark

Step 1: Collecting industry reference speeds
Industry reference speeds for the three product scenarios, simple, medium, and complex, were collected from Supplier A for three known machine designs, each with a different technical maximum speed. For each machine design, both best-in-class and typical reference speeds were provided.

Step 2: Estimating reference speeds for remaining machine designs

Since Supplier A could only provide reference speeds for three machine designs, the reference speeds for the remaining machine designs had to be estimated. This was done by calculating two adjustment ratios based on the data that were available.

The first ratio, the Industry Adjusted Ratio Typical to Best-In-Class (IARTB), reflects how typical speed relates to best-in-class speed for each product scenario i and machine design j, as shown in Equation 4.5:

$$IARTB_{i,j} = \frac{TypicalSpeed_{i,j}}{BestSpeed_{i,j}} \tag{4.5}$$

The second ratio, the Industry Adjustment Ratio Complex to Medium (IARCM), reflects how best-in-class speed for the complex scenario differs from the medium scenario for each machine design j, as shown in Equation 4.6:

$$IARCM_{c,m,j} = \frac{BestSpeed_{c,j}}{BestSpeed_{m,j}} \tag{4.6}$$

An average of each ratio was calculated across the three known machine designs, as presented in Table 4.3. The standard deviation for each ratio was small, approximately 2-3%, indicating that the ratios are consistent across machine designs and therefore reliable for estimation purposes.

These average ratios were then used to estimate the reference speeds for the remaining machine designs, where the best-in-class speed for simple and medium was set equal to the technical maximum speed of the machine since this reflected the same relation as for the provided machine speed data. These Equations 4.7, 4.8, 4.9 shows how the typical speeds for the remaining machine designs were calculated.

Table 4.3: Industry speed adjustment ratios by machine design and product scenario

Machine design	Machine	Simple (IARTB)	Medium (IARTB)	Complex (IARTB)	Complex (IARCM)
Machine design 1	F	0.54	0.49	0.44	0.91
Machine design 2	A, B	0.59	0.45	0.39	0.95
Machine design 3	—	0.56	0.40	0.38	0.89
Average ratio	C,D,E,G	0.56	0.45	0.40	0.92
Standard Deviation		0.024	0.021	0.032	0.031

$$v_{TypicalSpeed,s} = v_{BestSpeed,s} * Average(IARTB_s) \tag{4.7}$$

$$v_{TypicalSpeed,m} = v_{BestSpeed,m} * Average(IARTB_m) \tag{4.8}$$

$$v_{TypicalSpeed,c} = v_{TypicalSpeed,m} * Average(IARCM_{c,m}) \tag{4.9}$$

These Equations 4.10, 4.11, and 4.12 below presents the industry reference speeds after being adjusted for machine designs which in the next step were used to calculate the industry weighted average speed.

The industry reference speeds for each product scenario were denoted s = simple, m = medium, and c = complex, and performance level was denoted k, where k = best-in-class or typical, and machine designs were denoted j, where j = 1,2,3, or the average ratio, presented in Table 4.3.

$$v_{s,k,j} = speed_{Industry,s,k,j} \tag{4.10}$$

$$v_{m,k,j} = speed_{Industry,m,k,j} \tag{4.11}$$

$$v_{c,k,j} = speed_{Industry,c,k,j} \tag{4.12}$$

Step 3: Calculating the weighted industry speed for each plant

Once reference speeds were available for all machine designs j, a weighted average industry speed was calculated for each plant based on its product mix. This accounts for the fact that different plants produce different proportions of simple, medium, and complex products, and therefore a simple average would not reflect the reality of each plant's operational conditions. The weighted industry speed for both best-in-class and typical performance levels (k = best-in-class, typical) was calculated according to the Equation below for each machine A to G:

$$v_{WeightedIndustry,k,j} = w_s v_{s,k,j} + w_m v_{m,k,j} + w_c v_{c,k,j} \tag{4.13}$$

For comparison purposes, an unweighted average industry speed was also calculated for each machine A to G, by calculating an average of the reference speeds for simple, medium, and complex:

$$v_{AverageIndustry,k,j} = \frac{v_{s,k,j} + v_{m,k,j} + v_{c,k,j}}{3} \tag{4.14}$$

This enabled a comparison between the weighted and unweighted approaches to assess the impact of incorporating product-mix weights into the calculation. Figure 17 shows the result, where each machine's run speed is compared to its internal target and the weighted industry typical, best-in-class and technical maximum speed. As shown in the Figure 17, all machines except machine G perform above the typical industry speed. However, all machines fall below the best-in-class industry speed, indicating that there is remaining improvement potential for all plants. The dark band represents the range between best-in-class and technical maximum speed. This band is narrow, meaning that best-in-class performance in the industry is close to the technical maximum speed. It should be noted that the run speed bar shown represents each machine's average run speed and does not account for product mix.

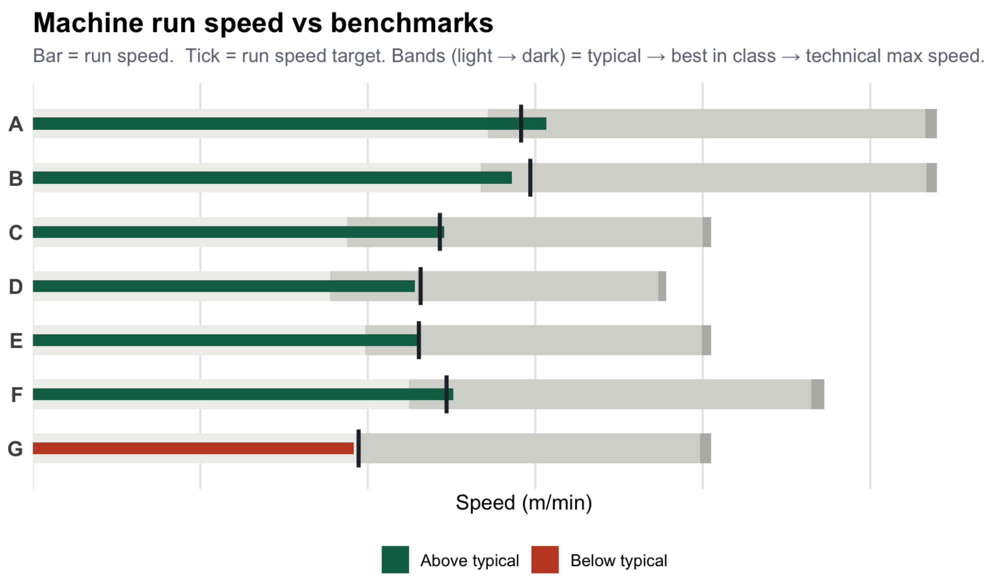


Figure 17: Machine run speed compared to weighted industry benchmark levels, showing typical, best-in-class and technical maximum speed

Figure 18 illustrates the difference between the weighted and unweighted industry benchmark speeds for best-in-class performance, where the weighted benchmark is shown in blue and the unweighted benchmark is shown in yellow. The weighted benchmark is calculated using each plant's actual product mix distribution according to Equation 4.13, while the unweighted benchmark represents a simple average of the three product scenarios as shown in Equation 4.14. As shown in Figure 18, the weighted benchmark speed is higher than the unweighted benchmark speed for all machines. This is explained by the fact that all plants run a higher proportion of simple and medium products. Simple and medium products, which have higher reference speeds than complex products. This means that the product mix pulls the weighted average upward. The same pattern is observed for the typical industry benchmark, as shown in Appendix B.1.

Figure 18: Weighted vs unweighted typical industry benchmark speed per machine

Figure 19 shows each machine's average run speed plotted against the typical reference speeds provided by Supplier A for each product scenario simple, medium, and complex, prior to any product-mix weighting. All machines have a higher average run speed than the reference speed for medium. Notably, machines C, D, and E have average run speeds that surpass the reference speed for simple products, and their internal speed targets are set above this level as well. Machine G, while still above the medium reference speed, shows the biggest difference between the simple industry reference speed and the internal speed target. As mentioned earlier, there is an important limitation in how speed targets are currently set. The case company sets its speed targets based on average run speed, without taking into account the product mix. Since no categorization of products into simple, medium and complex exists before this study, current targets do not reflect the actual complexity of the products being run. As seen in Figure 19, this leads to targets that can be misleading. For all the machines, the average run speed and run speed targets is higher than the typical industry reference speed for complex products. This means that speed targets are being overestimated for these products. At the same time, for simple products the speed target falls below the typical industry reference for some machines. This means that the speed targets are being underestimated and that there is room to run faster when producing simpler products

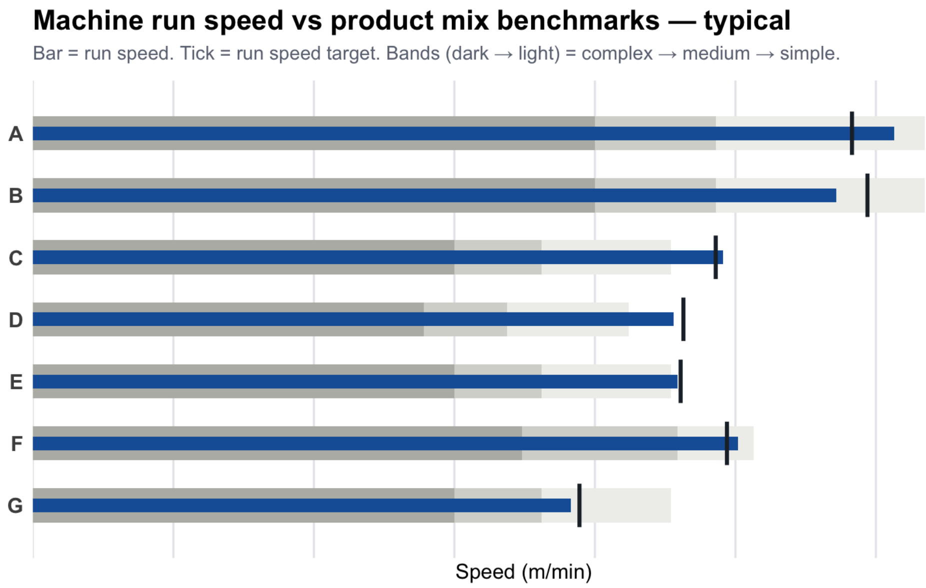


Figure 19: Machine run speed compared to product mix-specific typical industry benchmark speeds for simple, medium and complex products

4.3.2 Time Efficiency

Due to the design of the primary machine, setup time is recorded as zero, and the benchmarking of time efficiency therefore focuses on scheduled and unscheduled downtime only. The contextualization of the machine data in an industry context was calculated by taking the scheduled available time and applying the percentage which Supplier A stated should be dedicated to scheduled downtime. The scheduled and unscheduled downtime rate for each plant was calculated using the following Equations:

$$ScheduledDowntime_{rate} = \frac{ScheduledDowntime}{ScheduledAvailableTime} \tag{4.15}$$

$$UnscheduledDowntime_{rate} = \frac{UnscheduledDowntime}{ScheduledAvailableTime} \tag{4.16}$$

Figure 20 illustrates each machine's scheduled and unscheduled downtime rate compared to the industry benchmark for scheduled downtime recommended by Supplier A. As shown, all machines except machine A exceed the industry benchmark for scheduled downtime, suggesting that most machines dedicate more time to planned maintenance than what the supplier recommends as a minimum. Machine A is the only machine falling below the benchmark. Regarding unscheduled downtime, machines A and B show notably higher rates compared to the remaining machines, while machine F displays the lowest unscheduled downtime across all seven machines.

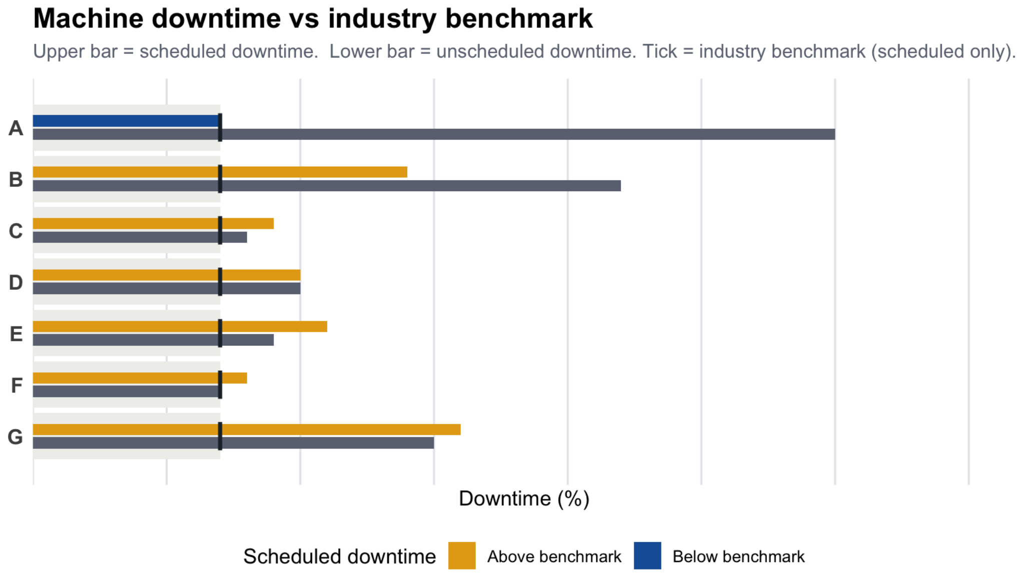


Figure 20: Scheduled and unscheduled downtime rates per machine compared to the industry-recommended scheduled downtime benchmark

As presented in the qualitative results, industry benchmark data for unscheduled downtime was not available from Supplier A, and technical unavailability was therefore used as a comparable measure instead. Since it could not be determined what proportion of scheduled or unscheduled downtime consists of technical unavailability, it was treated as a separate measure. Data for technical unavailability was available both from the supplier and in the company's internal data, enabling a direct comparison. Each machine's technical unavailability rate was calculated using Equation 4.17.

$$TechnicalUnavailability_{rate} = 1 - TechnicalAvailability \tag{4.17}$$

Figure 21 shows each machine's technical unavailability rate compared to the industry standard. As shown in figure 21 all machines exceed the industry standard expect machine E. Machines B and G shows the highest technical unavailability rates.

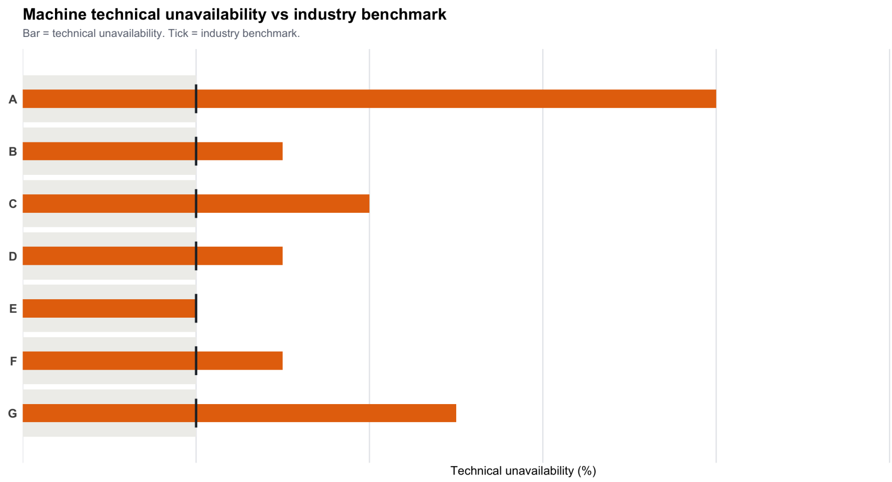


Figure 21: Machine technical unavailability vs industry benchmark

4.3.3 Material Efficiency

The material efficiency KPI, which is affected by Category 1 waste and Category 2 waste was divided into these categories and the waste rates for each machine were calculated and compared to the industry ratios for performance conditions best-in-class, typical, and poor conditions. This was done to enable a comparison of percentages to percentages. A full analysis of the material efficiency KPI was not possible, as this KPI is being measured plant level and not machine level. Instead, the wast rates for the two categories were calculated based on the input raw materials in tons and then compared to the industry levels of best-in-class, typical, and poor conditions. The waste rates was calculated using the following Equations:

$$Category1waste_{rate} = \frac{Category1waste}{InputMaterial} \tag{4.18}$$

$$Category2waste_{rate} = \frac{Category2waste}{InputMaterial} \tag{4.19}$$

$$TotalWaste_{rate} = \frac{Category1waste + Category2waste}{InputMaterial} \tag{4.20}$$

Figure 22 shows the total waste rate for each machine compared to the industry benchmarks. The majority of machines, A, B, C, D and F, have a total waste rate above best-in-class but below typical industry levels. Machines E and G exceed the typical industry waste rate. No machine achieves best in class waste performance.

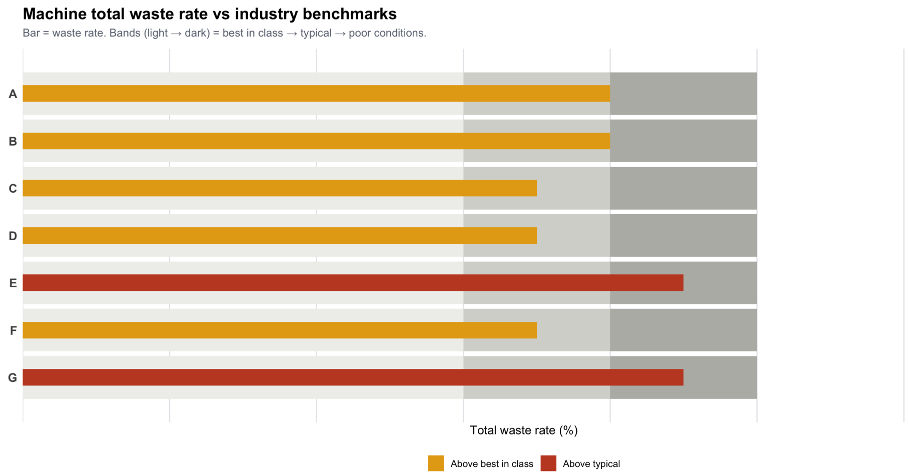


Figure 22: Machine total waste rate and industry rate comparison

Figure 23 shows the Category 1 waste rate for each machine compared to the industry benchmarks. Machines B, C, D, and F perform above the typical industry level, their waste rates fall between typical and poor conditions. Machines A, E, and G exceed the poor conditions benchmark, showing that their Category 1 waste rates are above what is considered poor performance in the industry.

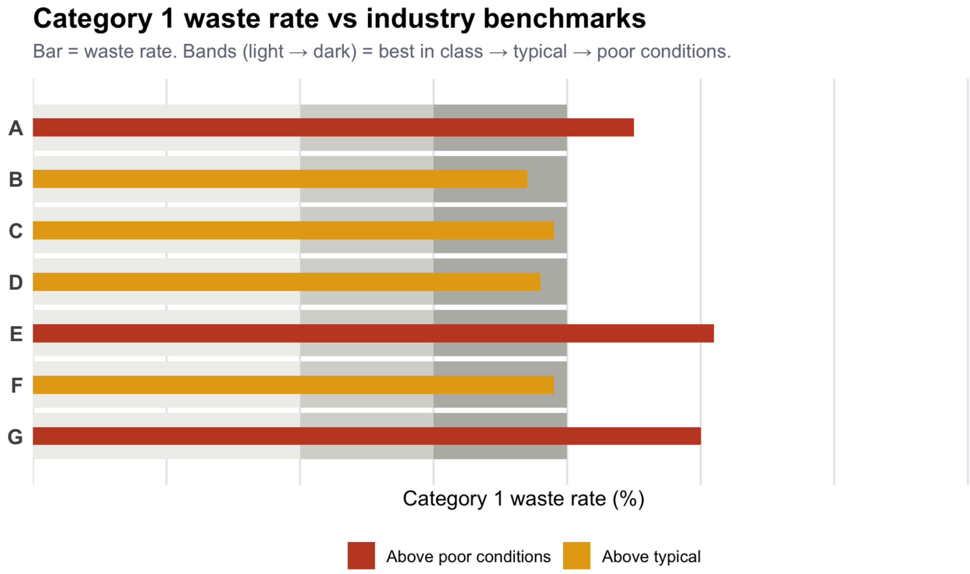


Figure 23: Category 1 waste rate and industry rate comparison

Figure 24 shows the Category 2 waste rate for each machine compared to the industry benchmarks. Machines A, C, D, and F have a waste rate below best-in-class, meaning their Category 2 waste rates are better than the best in class industry level. Machines B, E, and G fall between best-in-class and typical, meaning their Category 2 waste rates are within an acceptable industry range but have room for improvement towards best-in-class performance.

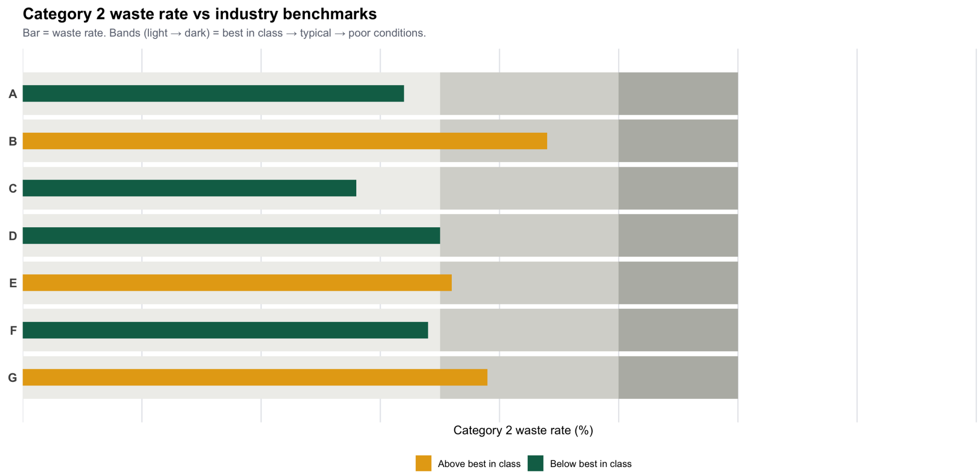


Figure 24: Category 2 waste rate and industry rate comparison

4.3.4 Industry Benchmark Performance Indices

The previous sections compared each machine's performance to industry reference levels for speed, time, and material efficiency. However, these comparisons do not clearly show how large the gap is between the machines and the industry level. To make this easier to understand, performance indices were calculated for each machine and KPI. An index value of 1 means the machine performs at the same level as the industry reference. Whether a value above or below 1 is favorable of unfavorable differs depending on the KPI, and is explained for each index below.

To calculate the speed efficiency index, the company's speed efficiency was first calculated by dividing the actual run speed by the technical maximum speed (Equation 4.21). The same approach was then applied to the industry reference speeds, giving an industry speed efficiency for both best-in-class and typical (Equation 4.22). The benchmark index was then calculated by dividing the company's speed efficiency by the industry speed efficiency, for both best-in-class and typical (Equation 4.23). An index value above 1 indicates that the machine performs above the industry reference level, while a value below 1 indicates a gap.

$$SE_{Company} = \frac{v_{act}}{v_{technicalmax}} \tag{4.21}$$

$$SE_{Industry,k} = \frac{v_{WeightedIndustry,k,j}}{v_{TechnicalMax}} \tag{4.22}$$

$$SE_{Bench,k} = \frac{SE_{Company}}{SE_{Industry,k}} \tag{4.23}$$

Figure 25 illustrates the speed efficiency index compared to best-in-class industry performance. As shown in Figure 25, all machines fall below the industry reference level of 1, with index values ranging from 0.48 to 0.61. This indicates that no machine currently performs at best-in-class level. However, the machines are relatively close to each other, suggesting that the performance gap to best-in-class is consistent across all plants rather than being specific to a few underperforming sites.

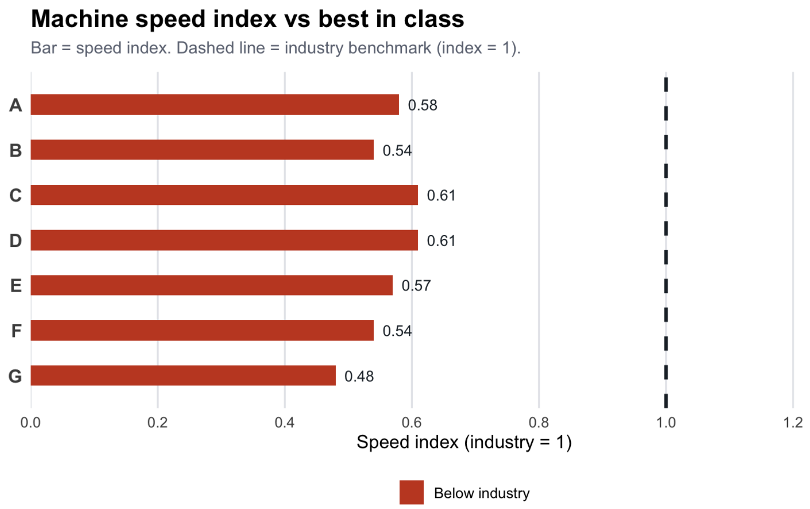


Figure 25: Speed Efficiency Index vs. Best-in-Class

Figure 26 illustrates the speed efficiency index compared to typical industry performance. As shown in Figure 26, all machines perform above or equal to the typical industry reference level of 1, with index values ranging from 1.0 to 1.3. This indicates that the company's machines are generally performing above what is typical in the industry in terms of speed efficiency. Machines C and D show the highest index values at 1.3, while machine G performs at exactly the industry level with an index of 1.0.

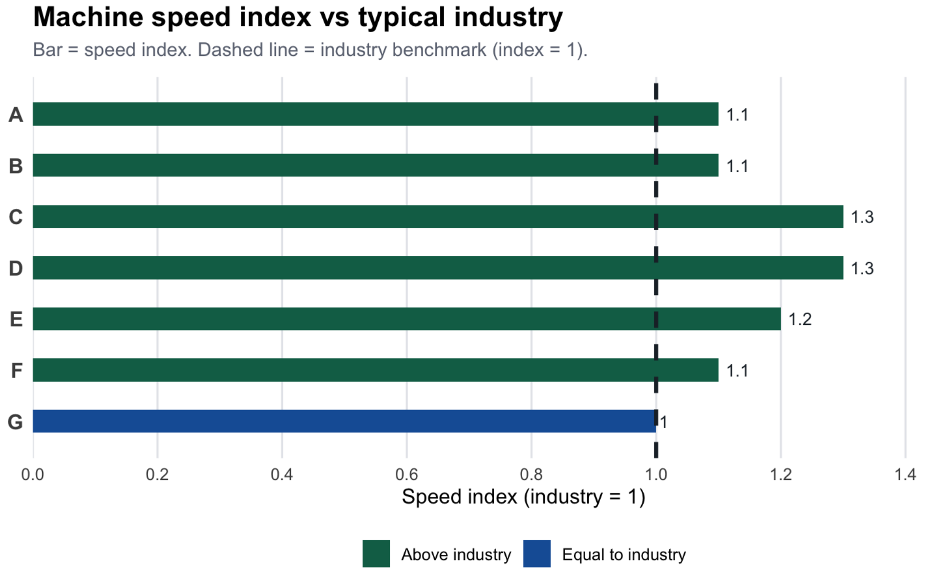


Figure 26: Speed Efficiency Index vs. Typical Industry Level

For time efficiency, two indices were calculated. The first compares each machine's scheduled downtime rate to the industry recommended level provided by Supplier A.

The second compares each machine's technical unavailability to the industry standard. As discussed in Section 4.2, a full benchmark of the time efficiency KPI was not possible since unscheduled downtime could not be benchmarked directly. The scheduled downtime index and technical unavailability index therefore provide a partial picture of each machine's time efficiency in an industry context.

The scheduled downtime index was calculated by dividing each machine's actual scheduled downtime rate by the industry recommended scheduled downtime rate, as shown in Equation 4.24.

$$SchedDT_{Bench} = \frac{SchedDT_{Company}}{SchedDT_{Industry}} \tag{4.24}$$

Figure 27 illustrates the scheduled downtime index compared to the industry recommended level. As shown in Figure 27, most machines have a scheduled downtime index above 1, meaning they dedicate more time to scheduled maintenance than the industry levels. Machine A is the only exception, with an index of 0.98, which can be considered equal to the industry level. The remaining machines range from 1.2 to 2.3, with machine G showing the highest index. As discussed previously, whether a higher or lower scheduled downtime is desirable depends on the perspective taken, and there is therefore no straightforward interpretation of this index.

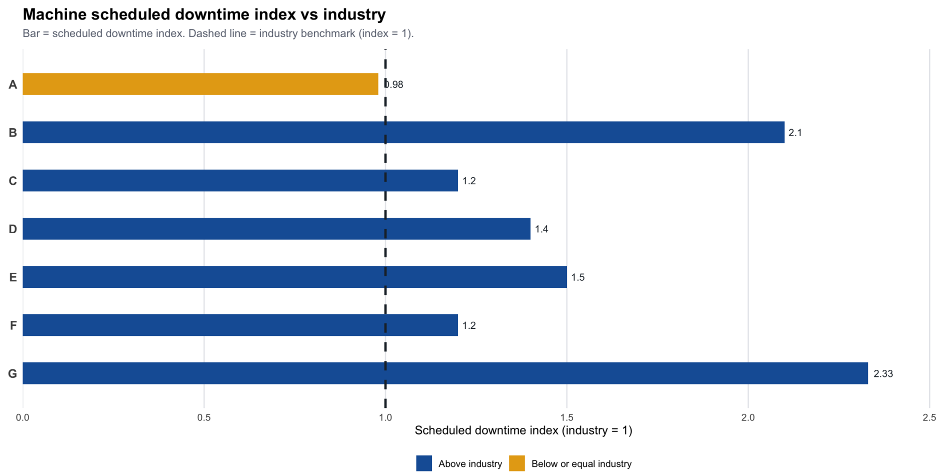


Figure 27: Scheduled Downtime Index vs. Industry Recommendation

The industry benchmark index for technical unavailability (TU) was calculated by dividing the company's TU by the industry standard TU, as seen in Equation 4.25, the best-in-class TU previously discussed by Supplier A was not taken into consideration in this case, since the industry standard is more representative.

$$TU_{Bench} = \frac{TU_{Company}}{TU_{Industry}} \tag{4.25}$$

Figure 28 illustrates the technical unavailability index compared to the industry standard. As shown in Figure 21, only machine E performs at or below the industry reference level with an index of 0.98, meaning it can be considered at industry level. All other machines exceed the reference level of 1. This means that most machines have a higher technical unavailability than the industry standard, which is undesirable since a lower technical unavailability indicates that the machine is available for production for a greater share of the planned production time. Machine A shows the largest gap with an index of 3.9, while machine B performs closest to the industry level among the underperforming machines with an index of 1.3.

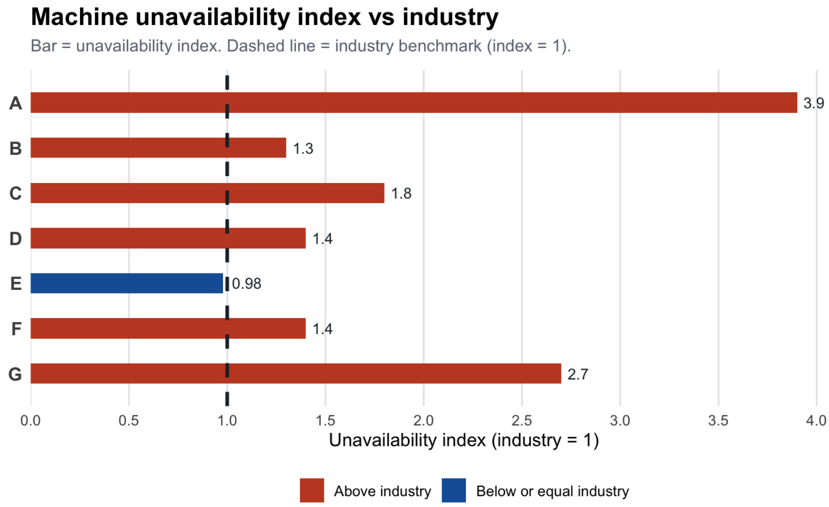


Figure 28: Technical Unavailability Index vs. Industry Standard

For the material waste the industry benchmark indices were calculated by dividing the company's waste ratio with the industry waste ratio for each category of waste respectively. Then, the total company waste ratio and the industry waste ratios was calculated to provide the complete picture of the waste industry comparison denoted h = best-in-class waste ratio, typical, and poor conditions waste ratio. Category 1 waste ratio was called (WR1) and Category 2 waste ratio was called (WR2) and the total waste ratio was called (WR). This can be seen in Equation 4.26,4.27 and Equation 4.28.

$$WR1_{Bench,h} = \frac{WR1_{Company}}{WR1_{Industry,h}} \tag{4.26}$$

$$WR2_{Bench,h} = \frac{WR2_{Company}}{WR2_{Industry,h}} \tag{4.27}$$

$$WR_{Bench,h} = \frac{WR_{Company}}{WR_{Industry,h}} \tag{4.28}$$

Figure 29 illustrates the total waste index compared to best-in-class industry performance. For this index, a value above 1 indicates that the machine produces more waste than the industry reference, which is a sign of worse performance. As shown in Figure 29, all machines have a waste index above 1, meaning that all machines generate more waste than the best-in-class industry level. Since a lower waste index is desirable, this indicates that there is improvement potential for all plants. The index values range from 1.2 to 1.6, with machines E and G showing the largest gap to best-in-class performance. As shown in Figure 23, the gap is primarily driven by Category 1 waste, meaning that reducing this type of waste would have the greatest impact on closing the gap to best-in-class performance.

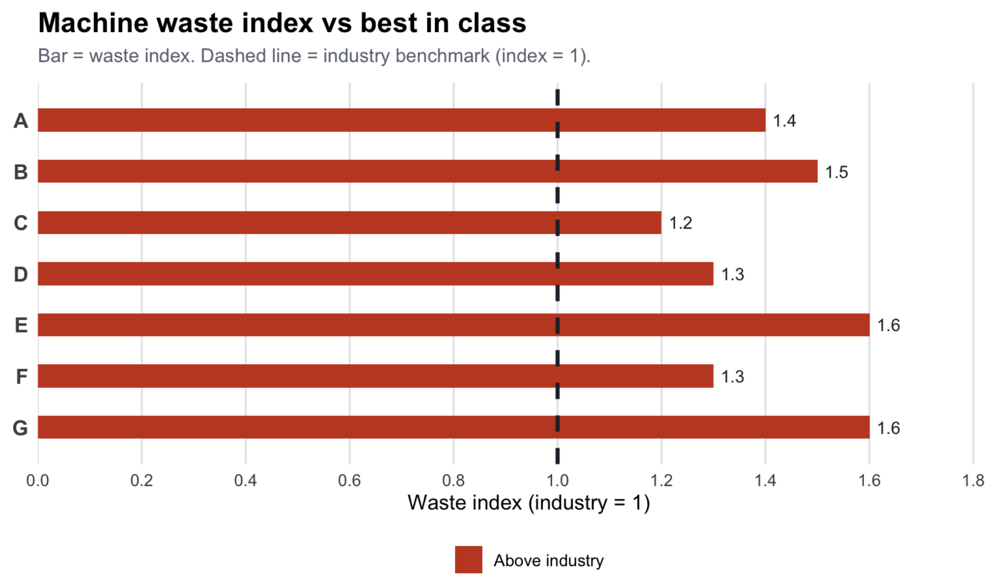


Figure 29: Waste Index vs. Best-in-Class Level

Figure 30 illustrates the total waste index compared to typical industry performance. As shown in Figure 30, most machines perform at or below the typical industry reference level, with index values ranging from 0.8 to 1.1. Machines A,B, C, D, and F all perform below or equal to the typical industry level, indicating that their waste levels are comparable to or better than what is typical in the industry. Machines E and G exceed the reference level with an index of 1.1, meaning that their waste levels are higher than what is typical. This is consistent with the results for Category 1 waste in Figure 23, where machines E and G also show the highest waste levels. Although machine A also has a higher Category 1 waste level than the typical industry reference, it performs better than machines E and G on Category 2 waste, as shown in Figure 24 which likely explains why machine A still falls below 1 in the total waste index for the typical industry reference.

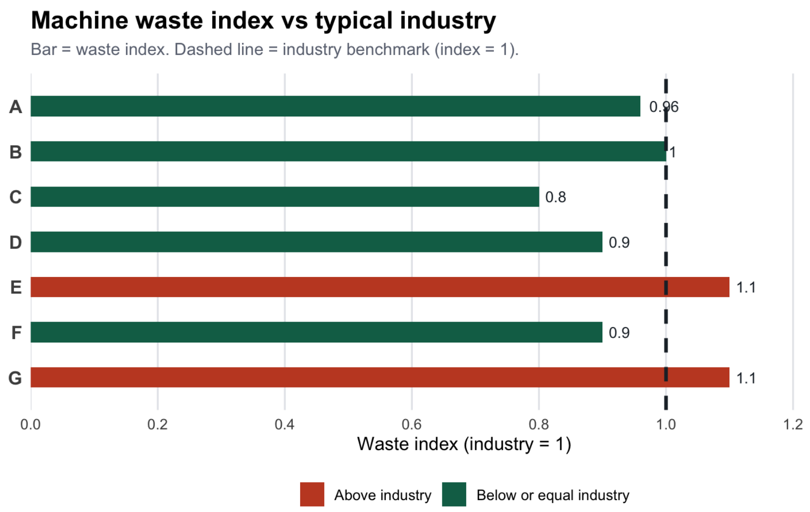


Figure 30: Waste Index vs. Typical Industry Level

Figure 31 illustrates the total waste index compared to poor industry conditions. As shown in Figure 31, all machines perform below the poor conditions reference level, with index values ranging from 0.7 to 0.9. This means that all machines generate less waste than what is considered poor performance in the industry, which is a positive result.

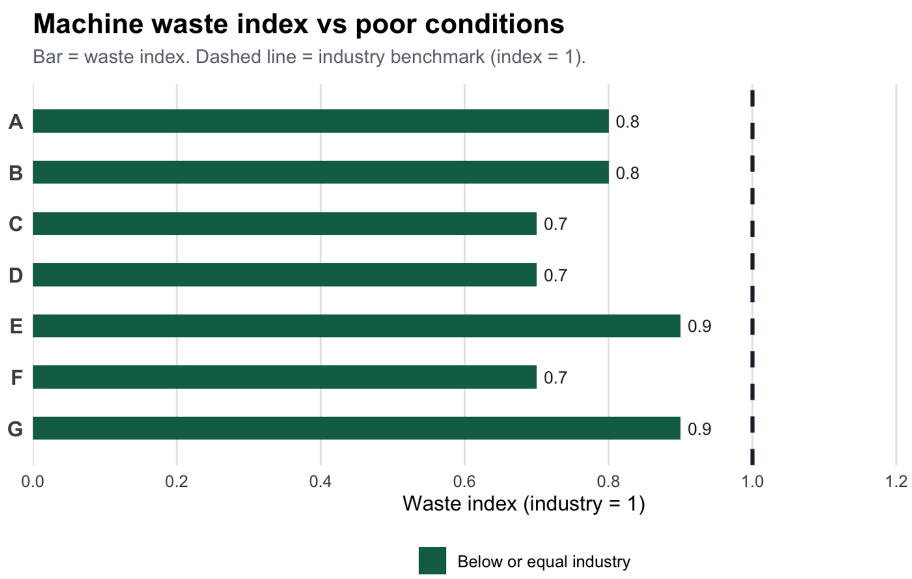


Figure 31: Waste Index vs. Poor Industry Conditions

5

Discussion

This chapter discusses the empirical findings in relation to the theoretical framework, structured around the two research questions.

5.1 RQ1: Which KPIs can be benchmarked externally?

The findings of this study demonstrate that external benchmarking can serve as a valuable comparative tool for increasing knowledge about current performance levels, which is consistent with Hong et al. (2012) who described benchmarking as essential for organizations seeking to understand their competitive position. Understanding one's competitive position is central to the Five Forces framework presented by Porter (1980), which argued that the intensity of competition within an industry is rooted in its underlying structure, and that a company must understand where it stands relative to that structure in order to identify its strengths and weaknesses. Chen et al. (2026) built on this by showing that firms which maintain a higher awareness of external competitive forces consistently achieve better performance outcomes, and that combining competitive strategy with strategic resource allocation outperforms isolated approaches. This suggests that the benchmarking conducted in this study goes beyond a purely operational exercise, and can be seen as a way for the case company to better understand its position within the competitive forces that shape its industry. At an operational level, this was reflected by Respondent I, who emphasized that challenging existing internal targets through an external perspective is essential for motivating continuous performance improvements within the organization. As Attiany (2014) argued, benchmarking establishes a baseline for continuous improvement activities. In addition, Ahmad and Dhafr (2002) stated that KPIs serve as performance indicators that gain additional value when internal performance values are compared against external reference levels, while also highlighting that evaluating several measures collectively is of high value to avoid misleading conclusions based on isolated indicators. The following subsections discuss each KPI of interest in relation to its ability to be benchmarked.

5.1.1 Speed Efficiency

Speed efficiency was identified as the KPI component that could be fully benchmarked externally. As Anvari et al. (2010) noted, OEE measures the relationship between the equipment's actual performance and its performance capabilities under optimal manufacturing conditions. Calandreli et al. (2025) further specified this for speed efficiency as the ratio of actual speed and ideal speed. Since speed efficiency is calculated as actual run speed divided by technical maximum speed, it is an objective and measurable ratio. The technical maximum speed is a known value determined by machine design. This makes it directly comparable across companies. Respondent K noted that each machine is designed for a specific maximum speed. Supplier A explained that the maximum speed is largely determined by the machine's design specifications. Since speed efficiency is so closely tied to machine capabilities, benchmarking through machine suppliers is particularly well suited for external benchmarking. As Chotayakul and Punyangarm (2026) noted, credible and comparable performance measurement requires reliable benchmarks, which can be estimated from equipment specifications when historical data are lacking. Machine suppliers have direct knowledge of these specifications across multiple installations in the industry. This makes them a reliable source for external benchmark data.

However, the qualitative findings also show that speed efficiency is influenced by product characteristics. Respondent M noted that speed efficiency depends on the type of product being produced. Respondent K identified product structure and material composition as key determinants. Greater thickness reduces the achievable speed, while heavier products generally result in lower run speeds. Respondent I further noted that product design alongside technical limitations influence speed efficiency. This introduces the challenge that Jonsson and Lesshammar (1999) highlighted. Comparing OEE across organizations is challenging as differences in operating conditions may lead to misleading results. For speed efficiency specifically, this challenge arise through the large variation of product mix. Supplier B highlighted that machine run speed is strongly influenced by each plant's production mix. Respondent I pointed out that benchmarking speed in an industry context requires consideration of several additional factors, including product mix, material composition, and order size. Without accounting for product mix, a plant running a higher share of complex products would appear to under-perform. This would not reflect actual inefficiency but rather differing operational conditions. Taking product mix into account when constructing the industry benchmark is therefore what makes speed efficiency a KPI that can be fully benchmarked. It resolves the comparability challenge that Jonsson and Lesshammar (1999) identified.

One dimension that is harder to benchmark externally is operator behavior and culture. Supplier A noted that cultural differences between countries can influence how operators run the machine. Respondent I highlighted that insufficient operator knowledge of when to increase run speed is one of the most significant constraints on performance improvement. As Alderman and Murray (2025) highlighted, a challenge with benchmarking is limited cooperation from relevant stakeholders. Since operator behavior and culture are largely dependent on human factors within each organization, these dimensions are difficult to capture through an external benchmarking approach with machine suppliers. Therefore, these factors should be kept in mind when interpreting the benchmark results, rather than being benchmarked directly.

5.1.2 Time Efficiency

Based on the results, it was not possible to fully benchmark the time efficiency metric through data collection from machine suppliers alone. This was primarily due to the fact that Supplier A could only provide information related to planned maintenance, and was unable to provide data on unscheduled downtime. A more complete benchmark could have been achieved if data had been collected directly from competitors or industry sources. However, according to Ofori-Kuragu (2025), selecting appropriate organizations as benchmarking peers can be challenging.

The collected data revealed that unscheduled downtime could not be obtained externally, as it was identified to depend on plant-specific conditions. The empirical results indicated that unscheduled downtime was primarily attributed to technical issues, required cleaning, lack of operator knowledge, and process-related disruptions, all of which are highly context-specific and therefore difficult to benchmark against a general industry reference level. Furthermore, Supplier A was unable to provide product-specific data on time efficiency, while Supplier B stated that time efficiency is not dependent on any product scenario specification. In contrast, Li et al. (2021) stated that in a multi-product OEE setting, products need to be treated separately in order to understand what drives performance. In addition, Trattner et al. (2019) outlined that product complexity affects processing time, and in this context Li et al. (2021) noted that different products cause bottleneck variability due to differences in processing settings. This suggests that at a product level, the supplier was unable to provide the downtime data needed to integrate the entire time efficiency KPI into the benchmarking initiative to compare with internal performance values.

Supplier A was able to provide recommended levels for scheduled downtime, enabling a holistic assessment of this KPI component. Calandreli et al. (2025) described scheduled downtime as one of the six production losses, while Supplier A stated that a certain level of scheduled downtime is necessary to extend the operational lifetime of the machines. This highlights a difference in perspectives between suppliers, manufacturing companies, and academic literature regarding whether scheduled downtime should be viewed as a production loss or as a long-term investment in machine health. Supplier A also pointed out the trade-off between maintenance cost and productivity, while highlighting an industry trend towards more hours being spent on scheduled downtime, indicating a broader shift towards preventive maintenance strategies.

Furthermore, Calandreli et al. (2025) included setup time within planned downtime, while the case company treats it as a separate component. This reinforces the argument made by Jonsson and Lesshammar (1999) regarding the difficulty of benchmarking OEE due to differences in definitions, which in turn limits the ability to make direct and meaningful comparisons across organizations. This is also supported by Cristea and Cristea (2021), who emphasized that KPIs vary across industries depending on the specific aspects that is needed to be measured.

Despite these definitional differences, neither Supplier A nor the case company included setup time in their definition of scheduled downtime, with both focusing instead on maintenance and cleaning activities aimed at preventing future machine issues. Their definitions can therefore be considered sufficiently aligned for comparison purposes, and the shared exclusion of setup time strengthens the validity of comparing their respective scheduled downtime values, mitigating the comparability concerns raised by Jonsson and Lesshammar (1999).

According to Calandreli et al. (2025), the two losses affecting availability, referred to by the case company as time efficiency, are scheduled downtime and unscheduled downtime. Although unscheduled downtime could not be benchmarked externally through machine suppliers, Supplier A provided data on industry standards for technical availability, indicating how well a production plant should be able to utilize the machine based on its technical conditions. This consideration is supported by Ahmad and Dhafr (2002), who suggested that achievable performance is influenced by process technology and the technical improvements made. Furthermore, Ullah et al. (2023) argued that KPI measurements can provide different signals depending on the perspective taken, whether from an internal or external view.

Furthermore, Ahmad and Dhafr (2002) pointed out that for the OEE KPI to reach high values, each of its components must also achieve high values. This strengthens the argument that even if a complete KPI comparison was not possible, partial indicators can still provide meaningful insights into how a machine performs relative to industry reference levels.

Based on the case company's definition of time efficiency and the external data provided by Supplier A, time efficiency could not be benchmarked as a complete KPI. However, the reference levels provided for scheduled downtime and technical availability still served as meaningful indicators of how the different machines performed relative to industry reference levels. This could in turn support informed target setting even where a full OEE comparison was not achievable.

5.1.3 Material Efficiency

Similar to time efficiency, the complete material efficiency KPI could not be benchmarked at a product scenario level, as Supplier A was unable to provide product scenario-specific waste data. This was explained by the fact that the material waste is not influenced by the type of product in the same way as speed, making product-specific waste data less meaningful in this context. This is in contrast to Li et al. (2021), who stated that in a multi-product OEE setting, products need to be treated separately in order to understand what drives performance and identify hidden costs, suggesting that product-specific waste data would nevertheless have been preferable for a more precise and comprehensive benchmarking comparison.

The measurement of material efficiency introduced an additional challenge, as it was assessed at a factory level rather than being isolated to the primary machine alone. As seen in Figure 8 in section 2.6.8, the total waste reflects the combined output of both the primary machine and the secondary machines, meaning that the primary machine represents only part of the total waste, creating irregularities when benchmarking the complete material efficiency KPI at machine level. Calandreli et al. (2025) described that OEE supports the identification of individual production losses, each of which can be separately connected to the three OEE components. This suggested that while a complete machine-level material efficiency KPI was not achievable, analyzing the individual waste categories separately could still provide meaningful insight into the primary machine's contribution to overall factory waste.

Supplier A was able to provide general industry reference levels for both Category 1 waste and Category 2 waste, enabling a holistic comparison across best-in-class, typical, and poor performance conditions. Despite this, the comparability of waste data across machines remains limited due to differences between the internal and external measurement systems. According to Supplier A, waste is measured differently across their customers, depending on whether it is captured directly at the machine or identified further downstream the process as poor quality output, strengthening the argument made by Jonsson and Lesshammar (1999) regarding comparability and definitions. Supplier A clarified that the industry waste levels provided were calculated based on rates from plants recording waste detected at later stages of the production process and caused by the primary machine. It was identified that the case company measures Category 2 waste as the output waste of the machine, without taking into consideration waste that is caused by the machine but detected later in the process. As noted by Jonsson and Lesshammar (1999), including downstream waste caused by the machine provides a more comprehensive view of the waste, although this can be difficult to achieve in practice. This suggests that the more effectively a plant can identify the sources of waste, the better it can understand where hidden costs arise.

This inconsistency in registering Category 2 waste affects the ability to benchmark performance and can lead to misleading conclusions. Nevertheless, even if Category 2 waste cannot be fully captured, it can still serve as a meaningful indicator of waste levels, and identifying the origin of detected waste losses can support improvement initiatives, which is consistent with Calandreli et al. (2025).

The reference levels provided by Supplier A for both Category 1 waste and Category 2 waste therefore still served as valuable indicators of how the case company's machines performed relative to industry waste levels. Even if a complete material efficiency comparison at a product scenario level was not achievable, the holistic reference levels provided could still be used to indicate how the industry performed in terms of waste generation and measurement practices. This provides a meaningful external reference point even in the absence of a complete material efficiency KPI.

5.2 RQ2: How can external benchmarking data be integrated with internal performance data?

As stated by Hong et al. (2012), benchmarking can be utilized to improve ways of working and achieve performance targets, starting with an evaluation of the company's current position. The benchmarking approach applied in this study was inspired by the first two phases of the Camp (1989) model, discussed by Anand and Kodali (2008) in their journal on benchmarking models. The steps that characterized this study were identifying the benchmarking subjects and partners, which consisted of the primary machine across the different production sites, the efficiency KPIs, and the machine suppliers. The data collection method was determined based on qualitative exploratory interviews with key personnel within the case company, with the aim of understanding how to approach the different KPIs in an external context. Following the data collection and analysis, the next step was to determine the current performance gap between the case company's machines and the identified industry reference level.

5.2.1 Selecting Suitable Peers

A challenge in benchmarking is identifying suitable peers to compare performance against. Alderman and Murray (2025) highlighted that finding a suitable reference point is a key challenge, as the choice of peer can limit comparability. The most direct approach would have been to benchmark against industry competitors, however, as Ofori-Kuragu (2025) pointed out, it is not always possible to find organizations willing to share their performance data when there are no mutual benefits. Drawing on the argument made by Porter (1980) that rivalry among existing competitors is a fundamental force shaping industry competition, it is reasonable that companies treat performance data as a competitive advantage rather than something to be shared with rivals. This was reflected in this study, where approaching direct competitors for performance data was not considered a realistic option given the confidential and competitively sensitive nature of production performance data. Supplier B was willing to share data, but only under strict conditions, primarily to avoid any risk of exposing or disadvantaging their existing customers. As Ofori-Kuragu (2025) pointed out, organizations are often reluctant to share performance data regardless of their role in the industry, as it can be seen as commercially sensitive. This explains why Supplier B was used primarily to confirm the product characteristics rather than as a data source.

In response to this challenge, Ofori-Kuragu (2025) argued that organizations should rely on alternative reference points that are realistically accessible. In this study, Supplier A was selected as the primary benchmarking peer, as they already had machines installed at the case company and were willing to collaborate without the same constraints as Supplier B. Furthermore, as they have machines installed across multiple companies in the industry, they possess both a high level of machine knowledge and a broad view of industry performance, making them a well-suited reference point.

However, using suppliers as benchmarking peers also comes with limitations. The data provided represents industry estimations rather than exact figures from competing plants. As Alderman and Murray (2025) noted, the choice of reference point can limit comparability, and this should be kept in mind when interpreting the results. That said, supplier-based benchmarking offered a practically accessible alternative that still allowed the case company to understand where it stood relative to the broader industry.

5.2.2 Product Specification

Before meaningful comparisons could be made across facilities, through the simulation, it was necessary to define a product specification capturing the key product characteristics that drive machine performance. However, no such product specification existed prior to this study, and developing one turned out to be a central part of the benchmarking process. The need for a product specification is closely linked to the nature of the production system. Li et al. (2021) identified product specification diversity as a key characteristic of multi-product production systems, where products made on the same line differ in physical attributes such as size, thickness and processing requirements, leading to different process settings and production speeds. This was clearly reflected in this study, where internal respondents highlighted that product characteristics such as grammage, thickness, width and structure all influence machine run speed differently. Li et al. (2021) also argued that using a single shared performance measure across all products will hide losses at the individual product level and give an inaccurate picture of overall performance. This meant that without first defining a product specification that captures the key characteristics driving performance differences, any comparison across facilities would risk being misleading.

To identify which characteristics were relevant, a relationship matrix was developed together with Supplier A, inspired by the House of Quality framework described by Erdil and Arani (2018). In this study, the relationship matrix was used to understand how each product characteristic influences the KPI components. This made it possible to identify not only which characteristics mattered, but also how strongly and in which direction they affected performance. The direction of each relationship was based on the concept of a ”sweet spot”, as described by Supplier A, referring to, for example, the grammage range where the machine performs at its best. Products within this range perform optimally, while products outside this range, either above or below, will cause performance to decrease. This is similar to how the relationship matrix in the House of Quality framework described by Erdil and Arani (2018) uses varying symbols to indicate both the strength and nature of relationships between requirements. In this study, the ”sweetspot” served as the basis for evaluating both the strength and direction of each relationship, providing a structured and consistent way of understanding how changes in product characteristics affect machine performance. Without this systematic approach, insights such as the interdependence of thickness and structure, or the negligible effect of width, would likely have gone undetected. This would risk producing an oversimplified and potentially misleading basis for comparison.

5.2.3 Categorize Products into Different Scenarios

Once the relevant product characteristics had been identified, the next step was to categorize products into different complexity scenarios: simple, medium, and complex. This categorization was a new addition for the case company, as no such structured way of grouping products based on complexity existed prior to this study. The categorization was defined by Supplier A, using the ”sweet spot” concept as the basis for distinguishing between simple, medium, and complex products. Products that fall within the ”sweet spot”, meaning the range where the machine performs at its best, were considered simple, as they require little or no speed reduction. Products outside this range were categorized as medium or complex depending on how far they differ from the optimal range and how much speed reduction they require. As Respondent I highlighted, establishing interval ranges for each product characteristic was important to make the categorization applicable across different production plants, since product characteristics are not standardized across facilities. This need for differentiation is supported by Trattner et al. (2019), who found that more complex products require more time and resources to produce than simpler ones. Without categorizing products by complexity, these inherent performance differences would be ignored, leading to comparisons that do not reflect the actual production conditions at each facility. Trattner et al. (2019) also argued that using physical product attributes that are specific to the industry gives the most useful basis for understanding how complexity affects performance, as more general measures of complexity tend to give less clear results. The results confirmed this approach, as products categorized as complex were found to require a heavier speed reduction than simple products, showing that industry-specific product attributes provide a more meaningful basis for understanding performance differences than generic measures of complexity.

5.2.4 Integration of External Benchmarking Data

In this section, the integration of the external benchmarking data was discussed in relation to theories focusing on the formulas and adjustments made for each KPI.

Speed Efficiency

When integrating the benchmarked speed values provided by Supplier A to enable a comparison with internal values, the technical maximum speed of each machine was an important factor to take into consideration. As stated by Respondent K and L, the technical maximum speed is dependent on the machine design, which introduced additional complexity into the integration calculations, as data was not provided for each individual machine type. This is consistent with Jonsson and Lesshammar (1999), who stated that optimal speed can vary depending on the type of products and machine design.

According to Anand and Kodali (2008), the data collection is typically defined during the planning phase of the benchmarking process. However, in this study, the process extended beyond data collection to include the adaptation of the data to enable fair and relevant comparisons within a simulation context. As a result, the data collection was conducted as an iterative process, where data was collected, analyzed, and redefined in multiple stages. This approach required the integration steps to account for product variability, as noted by Respondent M, as well as differences in technical maximum speeds across machine designs, in order to ensure comparability.

As previously discussed, the multi-product production system is highly dependent on the types of products being produced, which is consistent with Li et al. (2021), who argued that product-specific differences must be considered when evaluating performance KPIs. In contrast, a single-product production system would not need to account for product variations, since only one product contributes to the performance measured through KPIs. In such a case, the product mix would not need to be taken into consideration, simplifying the calculations considerably and making it easier to identify the sources of the six production losses as described by Calandreli et al. (2025). As stated by Li et al. (2021), producing only one product type reduces the number of variations to monitor, requires fewer adjustments in system settings and speeds, and simplifies the understanding of the overall system.

To account for the different machine designs, which Jonsson and Lesshammar (1999) and Respondent M emphasized as important, the IARTB and IARCM ratios were calculated to estimate industry reference speeds for machine designs where supplier data could not be directly provided due to variations in machine models. These ratios provided adjusted industry speeds for each machine design, which were subsequently converted into comparable values by multiplying them by the respective product-mix weights for each scenario. Relying solely on data from one machine type without adjusting for the technical maximum speed of each machine would not be representative of the broader industry context. This is consistent with Chotayakul and Punyangarm (2026), who stated that credible and comparable performance measurement requires reliable benchmarks, and Jonsson and Lesshammar (1999), who highlighted the importance of enabling meaningful comparisons of results through the use of a consistent reference speed.

According to Ahmad and Dhafr (2002), OEE levels in process industries are often below 50%, which implies that each of its components, speed efficiency, time efficiency, and material efficiency, would be expected to reflect correspondingly lower values. This can be further illustrated through the IARTB ratios, which indicate that the average typical industry speed is 56% of best-in-class speed or technical maximum speed looking at the simple product scenario. This indicates that the remaining efficiency measures, time efficiency and material efficiency, would each need to approach 100% for the OEE metric to reach 50%, which can be difficult to achieve in practice.

As identified in Figure 17 in Chapter 4, all machines exceeded the typical industry speed with the exception of Machine G, implying that the majority of the case company's machines perform at or above typical industry speed. According to Borrás et al. (2024), there is a need for setting appropriate reference points, and Ahmad and Dhafr (2002) stated that internal targets should be compared against external reference levels to ensure their relevance. In this context, the best-in-class speed may not represent a fully achievable or sustainable target over time. This is further supported by Supplier A, who noted that companies may approach speeds close to the technical maximum but cannot sustain them consistently. This is consistent with Ullah et al. (2023), who noted that internal performance that does not reach internal targets can still be strong relative to industry standards, suggesting that typical industry speed may serve as a more realistic and meaningful reference point, while a value between best-in-class and typical may represent a more relevant and achievable performance target.

Calandreli et al. (2025), discussed the six production losses and stated that two of the losses are connected to speed efficiency, which they refer to as performance losses. In the benchmarking initiative, the frequency of minor stops in the industry context was not taken into account, since these may be caused by unexpected events that are difficult to benchmark through machine suppliers. Furthermore, minor stops could potentially be connected to a lack of operator knowledge, as Respondent I mentioned that operator knowledge and confidence in operating the machine are among the largest constraints on machine performance.

The second loss connected to performance, reduced speed, as described by Calandreli et al. (2025), can be argued to have been considered in this study. The benchmarked run speed was adjusted based on the product being produced, which resulted in reduced speed relative to the speed that would have been used if it had not been taken into consideration, as seen in Figure B.1. Consequently, reduced speed in this context should not necessarily be interpreted as a performance loss, but rather as an inherent requirement of the specific product scenario, if the machine runs at the speed which is optimal for that scenario.

Furthermore, Jonsson and Lesshammar (1999) presented different calculations of speed efficiency, including the one proposed by Nakajima, in which the ideal cycle time is multiplied by the actual output and divided by the operating time. This can be interpreted as an approach suited to a single-product manufacturing system, while Puvanasvaran et al. (2013) described that for a multi-product production system, different cycle times for different products should be taken into consideration, representing their actual share of total production time. Differences in product characteristics can therefore, according to Li et al. (2021), increase organizational complexity and bottleneck variability, which highlights the difficulties in benchmarking, for instance, minor stops or setup times, as bottleneck variability is highly dependent on which products are being produced. This is consistent with Respondent I, who highlighted the product mix as an important factor to consider when benchmarking speed efficiency.

As mentioned, best-in-class speed is close to the technical maximum speed, and sustaining this level continuously is not realistic in practice. Ahmad and Dhafr (2002) noted that achievable performance targets are largely influenced by process technology, suggesting that older and less automated machines may not be able to reach the same speeds as newer machines. This is supported by Supplier A, who pointed out that older machines were designed for heavier products, making them less suited for the lighter products that are increasingly common in the industry today. As Respondent K noted, best-in-class performance further depends on the ability to quickly identify and solve unexpected problems, supported by advanced monitoring systems. Together, these factors help to explain why best-in-class speed may not represent a realistic or achievable target for all machines.

A key choice in this study was whether to apply product-mix weights when calculating the industry benchmark speed, or to use a simple average of the three product scenarios. As highlighted by Greco et al. (2019), giving equal weight to all criteria is not justified, as it ignores the difference between more and less important factors. Puvanasvaran et al. (2013) argued that a weighted average approach better accounts for differences in product mix, ensuring that products contribute proportionally to their actual share of total production time. In this study, the weights were based on each plant's actual share of run time dedicated to simple, medium, and complex products, inspired by their equation. As shown in Figure 18, the weighted benchmark speed is higher than the unweighted speed for all machines. This is explained by the fact that all plants run a higher proportion of simple and medium products, which have higher reference speeds than complex products, pulling the weighted average upward. The exact size of the difference varies across plants and has not been precisely quantified in this study. However, the results confirmed the arguments made by both Greco et al. (2019) and Puvanasvaran et al. (2013), suggesting that incorporating product-mix weights gave a more representative industry benchmark than a simple average would have done.

The indices presented in section 4.11 were calculated by dividing the company's performance value by the industry reference value, following the approach described by Oral et al. (1999). According to them, this type of relative performance measure provides valuable information about a firm's competitiveness by placing internal performance in an external context. An index value of 1 indicates performance at the industry reference level. A value above 1 indicates that the company performs above the industry level, while a value below 1 indicates a gap to the industry. For speed efficiency specifically, the index follows this logic. However, unlike the general case described by Oral et al. (1999), whether a value above or below 1 is desirable differs depending on the KPI. This is therefore clarified for each KPI index discussed. **Time efficiency**
As stated by Calandreli et al. (2025), the two production losses connected to availability, referred to in this study as time efficiency, are scheduled downtime and unscheduled downtime. As discussed above, unscheduled downtime could not be benchmarked externally, while scheduled downtime could.

As identified by Jonsson and Lesshammar (1999), formulas for time efficiency are presented differently in the literature, with different components taken into consideration depending on the definition applied. The formula used to integrate the external benchmarking values with the internal performance data was designed to account for the two losses connected to time efficiency as presented by Calandreli et al. (2025), namely scheduled downtime and unplanned downtime.

When integrating the benchmarked values into the internal context to simulate how the machines performance would be at industry reference levels, the scheduled downtime percentages provided by Supplier A were multiplied by the scheduled available time for each machine. Since the scheduled available time is a machine-specific value, the result of this multiplication indicates how each machine would perform if it operated at industry reference levels for scheduled downtime. However, since the unscheduled downtime percentage could not be obtained externally, the complete time efficiency KPI could not be fully benchmarked.

Had the unscheduled downtime value been available, it would also have been multiplied by the scheduled available time to determine the number of hours dedicated to unplanned downtime in a year. This would have enabled the calculation of the production time by subtracting both the industry unscheduled downtime and the industry scheduled downtime from the scheduled available time. Subsequently, to obtain the run time representing how each machine would perform at industry reference levels, the setup time for each machine would have been subtracted from the production time, which in the majority of cases were zero. This industry run time would then have been divided by the scheduled available time to obtain the time efficiency value at industry reference levels.

By applying this approach, the benchmarking initiative would simulate and provide values indicating how each machine would perform in terms of time efficiency when operating at industry reference levels, making the comparison more meaningful and adjusted to each machine's specific scheduled available time. This would enable the establishment of a suitable industry reference point, which, according to Alderman and Murray (2025) could serve as a baseline for continuous improvement activities. However, since the unscheduled downtime was not available, the company's percentage values for scheduled and unscheduled downtime were calculated to be compared with the percentage rates provided by the supplier.

When comparing the industry reference level of scheduled downtime against the actual scheduled and unscheduled downtime of the different machines, it can be identified that almost all machines have a scheduled downtime rate higher than the industry reference level, which may indicate that these machines require more hours dedicated to maintenance or cleaning. It can also be identified that machine A and B have a notably higher percentage of unscheduled downtime relative to their total downtime. According to Supplier A, dedicating more time to preventive maintenance could reduce unscheduled downtime over time. This is supported by Ahmad and Dhafr (2002), who stated that improvement approaches such as predictive maintenance and improved process control could result in higher overall performance. In contrast to machines A and B, the other machines C, E, F, and G all have a higher level of scheduled downtime than unscheduled downtime. As highlighted by Supplier A, the effectiveness of scheduled downtime ultimately depends on how it is utilized, and whether the maintenance activities performed are sufficient to minimize unscheduled downtime in the long run.

According to Ahmad and Dhafr (2002), technical improvements can improve the OEE components. Supplier A provided data on technical availability and technical unavailability, for which the industry standard level was provided as a percentage value, as seen in Figure 21. As shown, all machines have a higher technical unavailability than the industry standard, indicating that they are less technically available than what Supplier A considers to be an industry standard level.

For the scheduled downtime index, most machines perform above the industry recommended level, as shown in Figure 27. Only machine A performs at approximately the industry level with an index of 0.98. Unlike the general logic described by Oral et al. (1999), where an index above 1 indicates better performance, the scheduled downtime index does not follow this straightforward interpretation. As discussed earlier, there is a trade-off between maintenance and production time. A higher index, meaning more scheduled downtime than the industry recommendation, can be positive as it allows for preventive maintenance and can reduce unexpected downtime. However, from a production perspective, a lower index is preferable as it maximizes productive machine time. There is therefore no single correct interpretation, and each plant must decide what level is appropriate based on their own priorities. For technical unavailability, the index follows the relative performance logic described by Oral et al. (1999), however, with an inverted interpretation. Since technical unavailability is a measure where lower values indicate better performance, an index below 1 suggests that the company performs better than the industry reference, while an index above 1 indicates that the machine has a higher technical unavailability than the industry standard, which is undesirable.

Material efficiency

Jonsson and Lesshammar (1999) highlighted the different ways in which the quality component of OEE is defined in literature, with reference to Nakajima and De Grote, who estimated quality, referred to in this study as material efficiency, in similar but not identical ways. Nakajima calculated good output as the input minus the defective output divided by the total input, while De Groote defined quality as the ratio of accepted production relative to total production. The similarities between these definitions suggests that the underlying principle of measuring material efficiency as the proportion of usable output relative to total input is relatively consistent in literature, which provided a foundation for how waste ratios were calculated in this study.

For the benchmarking initiative, the material waste ratios were calculated by determining the waste ratio for each Waste Category relative to the case company's total input material in production. These ratios were then compared against the industry waste rates provided by supplier A for best-in-class, typical, and poor performance conditions. By expressing both internal and external values as ratios of input material, the values were made comparable across machines, as seen in Figures 22, 23, and 24.

The identified differences in measuring waste affect the comparability of the internal and external values. However, if all values were measured in the same way and material efficiency was calculated at machine level a different approach would have been taken during the simulation. The waste rates provided by the supplier would have been multiplied by the input material to obtain the industry waste measured in tons instead of percentages, the sum of the two waste categories would then be divided by the total input material as Jonsson and Lesshammar (1999) described Nakajima's quality KPI to be calculated.

For material efficiency, three waste indices were calculated, comparing each machine to best-in-class, typical, and poor industry conditions. The index is calculated by dividing the company's waste level by the industry waste level, following the approach of Oral et al. (1999). Similar to the technical unavailability index, waste is a measure where less is better, and the index therefore follows the same inverted logic, where an index below 1 is desirable and an index above 1 is undesirable.

5.3 Theoretical Implication

This study contributes to existing research in several concrete ways. Jonsson and Lesshammar (1999) highlighted that comparing OEE across organizations is problematic because companies define and measure performance differently. This study encountered and confirmed that challenge. Making internal and external data comparable required significant adjustments across sites. By developing product-specific complexity scenarios and weighting them according to each machine's actual production mix, this study extends their work by offering a practical solution. Adjusting external reference values to account for product complexity ensures that comparisons reflect actual operating conditions, rather than assuming all machines run the same types of products.

A key theoretical contribution relates to Li et al. (2021). They argued that products in a multi-product system must be treated separately to avoid misleading performance conclusions, a principle they developed in the context of internal production analysis. This study confirmed this, as there was no product specification prior to this study and developing one turned out to be necessary to make comparison possible. This also showed that the same principle applies when benchmarking against external reference values, extending their framework into a new setting.

Trattner et al. (2019) showed that more complex products require more time and resources to produce, and that industry-specific product attributes provide the most meaningful basis for capturing performance differences. This study confirmed both findings, as complex products required greater speed reductions than simple ones, and the industry-specific attributes used to define the complexity scenarios successfully captured these differences. This study also placed these findings in a new context, which showed that product complexity should be accounted for in external benchmarking, otherwise comparisons between machines become unfair.

Finally, the study showed that machine suppliers can work well as benchmarking partners when competitor data is not available. Ofori-Kuragu (2025) argued that benchmarking should be possible even without cooperation from top-performing competitors. This study provided a concrete example of how that can work in practice. Because suppliers operate across multiple production sites, they have both the technical knowledge and the broader performance overview needed to serve as a meaningful reference point.

5.4 Practical Recommendations

Based on the benchmarking results, introducing product-specific speed targets could be a valuable first step for the case company. As Li et al. (2021) argued, applying a single shared performance measure across all products in a multi-product system tends to hide losses at the individual product level, giving a misleading picture of overall performance. This is evident in Figure 19, where a single average target fails to distinguish between simple, medium, and complex products. Given that targets are displayed on dashboards next to machines for operators to follow, as noted by Respondent M, they play a direct role in shaping operational decisions. One operator observed during a Gemba walk that increasing run speed on complex products can lead to higher waste levels, suggesting that a uniform target may push operators to run complex products at speeds that risk disrupting quality, while simultaneously underutilizing the potential of simpler products. Jonsson and Lesshammar (1999) emphasized that the same speed reference must be applied consistently for a given product across all measurements. Without this, performance comparisons over time become unreliable. Product-specific targets would therefore not only improve operational guidance but also establish a more consistent basis for measurement. As Respondent I pointed out, production plants can carry internal biases about their own capabilities, and targets grounded in external industry benchmarks could help challenge these assumptions and direct performance improvement in a more informed way.

Beyond target-setting, there are behavioral and cultural factors that constrain speed efficiency, and these are closely connected to the broader challenge of reducing downtime. Supplier A noted that operators in some contexts tend to run machines at a comfortable pace rather than pushing toward the maximum achievable speed, and that middle management plays a critical role in addressing this, since operational issues are largely handled at that level. Respondent I similarly identified insufficient operator knowledge of when to increase run speed as one of the most significant performance constraints, and Respondent K reinforced this by pointing out that unscheduled downtime can also be reduced through improved operator knowledge and effective cross-team communication.

Ahmad and Dhafr (2002) suggested that OEE improvements are often achievable through operator training and better operating procedures, and this applies across both speed and time efficiency. Reducing downtime further requires a combination of technological and human investment. Respondent K noted that plants with advanced monitoring systems, such as cameras and sensors, combined with skilled operators, are better positioned to detect and resolve problems quickly. Supplier A highlighted that increasing scheduled downtime can reduce unscheduled stops by allowing potential issues to be identified and resolved proactively, a shift that Ahmad and Dhafr (2002) similarly connected to predictive maintenance and improved process control as pathways to higher overall performance. Antony et al. (2026) reinforced this by arguing that technological improvements are insufficient on their own, and that strong leadership, a capable workforce, and a supportive organizational culture are equally necessary for sustained performance gains.

On waste, the benchmarking results suggested that the company performed around the typical industry level, though the best-in-class comparison revealed room for improvement. Category 1 waste is closely tied to production planning, which Respondent M noted is largely driven by customer order patterns, meaning the focus should be on optimizing planning within existing constraints. For Category 2 waste, Supplier A highlighted that defects originating from the primary machine can travel downstream and be registered as other types of waste, distorting the statistics. Earlier detection through improved monitoring systems could both reduce waste and improve the accuracy of how it is recorded, which aligns with Antony et al. (2026), who highlighted real time monitoring as a key enabler of operational excellence. Ahmad and Dhafr (2002) further suggested that quality improvements are achievable by minimizing rework and process adjustments during production. Furthermore, if material efficiency were to be measured at the machine level rather than the plant level, this could enable more targeted identification of each machine's contribution to overall waste. As Calandreli et al. (2025) noted, OEE supports identification of production losses which in turn helps direct improvement efforts where they matter most. This aligns with Antony et al. (2026), who described operational excellence as rooted in principles from Lean Thinking, where waste reduction is a core lever for improving both productivity and cost efficiency.

Taken together, the benchmarking results across all three KPI components offer more than operational insight, they provide a basis for strategic positioning. Porter (1980) argued that understanding a firm's relative strengths and weaknesses is essential for identifying where action can make the greatest difference, and Baird et al. (2024) showed that benchmarking allows organizations to evaluate and reassess their competitive standing in response to the forces shaping their industry. In this sense, the gaps identified across speed efficiency, time efficiency, and material efficiency should be read not only as operational targets, but as indicators of where the company stands and where it has the most to gain. The approach applied in this study could also be extended to other machine types within the company, which could deepen the understanding of overall production performance and reveal where improvement efforts would have the greatest impact across the production chain. For such a comparison to be meaningful, two aspects are worth considering. The first is whether machines can be benchmarked under comparable circumstances, meaning whether a suitable external reference point exists with sufficient knowledge of the machine type and its industry context. The second is whether the product mix introduces complexity that affects machine performance. Where a machine produces a wide range of products with characteristics that influence how it operates, defining product-specific scenarios and corresponding weights becomes a necessary step before any fair comparison can be made. As Jonsson and Lesshammar (1999) stressed, meaningful performance comparisons require consistent and well-defined reference points, and this holds regardless of which machine type the approach is applied to.

The benchmarking initiative was presented to the case company upon completion of this study. The findings were received positively. Representatives highlighted the value of being able to visualize performance across different KPIs in relation to industry reference levels. This provided a clearer picture of where the company stands in an external context. Beyond the benchmarking results themselves, insights such as the value of introducing product-specific speed targets, the finding that waste caused by the primary machine but detected further downstream in the process is currently not captured in the company's waste measurement, and the potential of measuring material efficiency at the machine level rather than the plant level were also highlighted as valuable takeaways. The case company expressed interest in continuing to explore how this approach can be further developed and applied to their secondary machines.

6

Conclusion

This chapter presents the conclusions drawn from the two research questions, followed by a discussion on the study's limitations and directions for future work.

6.1 RQ1: Which KPIs can be benchmarked externally?

Based on the discussion of the external benchmarking of speed efficiency, time efficiency and material efficiency, it was possible to benchmark speed efficiency by taking each production plant's product mix into consideration. However, neither time efficiency nor material efficiency could be fully benchmarked at a product scenario level, nor could the complete KPIs be compared against external reference levels.

The main reasons identified as limiting the benchmarking of time efficiency and material efficiency were as follows: for time efficiency, the supplier was unable to provide data on unplanned downtime, while for material efficiency, the primary limitations were that it is measured at a more aggregated factory level and that inconsistencies in measuring Category 2 waste across the industry were identified. Instead, individual components of each KPI could be compared, providing an indication of how the case company's primary machines performed relative to different industry reference levels.

6.2 RQ2: How can external benchmarking data be integrated with internal performance data?

The external benchmarking data could be integrated with internal performance data through a simulation-based approach, where external benchmarking data were collected via machine suppliers and embedded into the company's own operational context. However, the method varied depending on whether a KPI could be fully benchmarked. For Speed Efficiency, which was fully benchmarked, the integration followed several steps:

1. Product characteristics affecting speed efficiency were defined and categorized into complexity scenarios (simple, medium, complex) based on how they affected machine runnability.

2. Each scenario was weighted by its share of total production run time to account for differences in product mix.

3. Supplier-provided speed data were adjusted per machine design to allow comparability, and multiplied by these weights to simulate industry-level speeds within the company's own operational context.

4. A benchmarking index was calculated as a ratio between the actual run speed and the simulated integrated industry run speed.

For the time efficiency and material efficiency KPIs, the entire KPI could not be benchmarked. Instead, individual components of the KPI could be directly compared against supplier-provided percentages to calculate a component-level index. Since the supplier stated that product characteristics did not have as great an impact on time efficiency and material efficiency as on speed efficiency, product scenarios were not included as a step in the benchmarking approach for these KPIs. To achieve a full integration, the entire KPI would need to be benchmarked, meaning that the component percentages provided by the supplier would need to be multiplied by scheduled available time for time efficiency and by input material for material efficiency. A further integration approach would continue with the company's values for scheduled available time and input material as baseline inputs, while replacing the company-specific variables in the KPI formulas with values derived by applying industry-based percentages to these baseline values. This would enable the calculation of fully benchmarked KPIs.

The core value of this benchmarking approach lay in its ability to simulate how each machine would perform under industry conditions within its own operational context, rather than comparing against unadjusted industry benchmarks. By translating supplier-provided data into the company's operational context, the approach enabled a fair and machine-specific assessment of performance relative to industry reference levels. This allowed performance gaps to be identified and further improvement opportunities to be discovered.

6.3 Limitations and Future Work

The scope of this research, which involved investigating the performance of seven machines across seven different plants, necessitated a holistic approach. Consequently, the product scenario definitions were developed based on common characteristics identified during the interviews. As a result the opportunity to conduct a more detailed analysis of the underlying data and assess the extent to which it supported the categorization was limited. The product-mix weights were also subject to limitations, as they were obtained from secondary sources and were therefore not directly analyzed by the authors. Furthermore, the industry reference data were collected through single machine supplier, which may limit the generalizability of the results. However, the supplier was still able to provide valuable industry insights due to its extensive experience and broad expertise within the field.

Further research topics identified in this study involve, firstly, analyzing and investigating industry performance in relation to internal performance over several years in order to examine how machine performance and industry reference levels evolve over time. Secondly, future research could investigate how a more in-depth categorization of products in a multi-product production system, based on their impact in machine performance, could be conducted. This study adopted a more holistic approach to enable benchmarking, however, a more detailed categorization could provide value for internal OEE calculations and the identification of hidden costs. Thirdly, further research could investigate what would be required to benchmark time efficiency and material efficiency at a product-level. Fourthly, future research could be extended to the secondary machine in the production process. However, it should be noted that different product characteristics may need to be identified, as the secondary machine represents an additional step further into the production process with potentially different performance drivers.

References

Ahmad, M. M., & Dhafr, N. (2002). Establishing and improving manufacturing performance measures. *Robotics and Computer-Integrated Manufacturing*, *18*, 171–176.

Alderman, L., & Murray, S. (2025). Benchmarking: Seeking best practice. *New Directions for Evaluation*, (185–186), 25–32. https://doi.org/10.1002/ev.20634

Allott, J., Grünewald, F., & Luyeye, L. (2023, May). *Responding to productivity improvement challenges in pulp and paper* (Industry Report). McKinsey & Company. Retrieved 2025, from https://www.mckinsey.com/industries/paper-forest-products-and-packaging/our-insights/responding-to-productivity-improvement-challenges-in-pulp-and-paper

Anand, G., & Kodali, R. (2008). Benchmarking the benchmarking models. *Benchmarking: An International Journal*, *15*(3), 257–291. https://doi.org/10.1108/14635770810876593

Antony, J., Kaul, A., Bhat, S., Sony, M., Yamini, S., Citybabu, G., & Cudney, E. A. (2026). Operational excellence themes: Past, present and future. *The TQM Journal*. https://doi.org/10.1108/TQM-07-2025-0410

Anvari, F., Edwards, R., & Starr, A. (2010). Evaluation of overall equipment effectiveness based on market. *Journal of Quality in Maintenance Engineering*, *16*(3), 256–270. https://doi.org/10.1108/13552511011072907

Attiany, M. S. (2014). Competitive advantage through benchmarking: Field study of industrial companies listed in amman stock exchange. *Journal of Business Studies Quarterly*, *5*(4), 12–?

Baird, K., Nuhu, N., & Jiao, L. (2024). The effect of Porter's competitive forces on competitive advantage and organisational performance and the moderating role of management accounting practices. *Journal of Management Control*, *35*, 303–332. https://doi.org/10.1007/s00187-024-00375-4

Borrás, F., Ruiz, J. L., & Sirvent, I. (2024). Planning improvements through data envelopment analysis (dea) benchmarking based on a selection of peers. *Socio-Economic Planning Sciences*, *95*, 102020. https://doi.org/10.1016/j.seps.2024.102020

Calandreli, P. R., Valle, P. D., & Deschamps, F. (2025). Maximizing operational efficiency with industry 4.0 technology: Integrating oee as a performance indicator. *The International Journal of Advanced Manufacturing Technology*, *138*, 855–872. https://doi.org/10.1007/s00170-025-15408-y

Chen, Y.-M., Kao, Y.-F., Yan, M.-R., & Yen, M.-H. (2026). Managing competitive advantage in emerging knowledge-based economies: The strategic configuration perspective. *Knowledge Management Research & Practice*. https://doi.org/10.1080/14778238.2026.2615351

Chotayakul, S., & Punyangarm, V. (2026). Green overall equipment effectiveness (goee): Theoretical development and simulation-based analysis for sustainable manufacturing. *Cleaner Engineering and Technology*, *31*, 101182. https://doi.org/10.1016/j.clet.2026.101182

Cristea, C., & Cristea, M. (2021). Kpis for operational performance assessment in flexible packaging industry. *Sustainability*, *13*, 3498. https://doi.org/10.3390/su13063498

Edwards, W., & Barron, F. H. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, *60*(3), 306–325. https://doi.org/10.1006/obhd.1994.1087

Erdil, N. O., & Arani, O. M. (2018). Quality function deployment: More than a design tool. *International Journal of Quality & Service Sciences*, *11*(2), 142–156.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking qualitative rigor in inductive research: Notes on the gioia methodology. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, *141*, 61–94. https://doi.org/10.1007/s11205-017-1832-9

Hammond, M., & Wellington, J. (2021). *Research methods: The key concepts* (2nd ed.). Routledge.

Hardin, M., Horn, D., Perez, R., & Williams, L. (2012, January). *Which chart or graph is right for you? tell impactful stories with data* (tech. rep.). Tableau Software.

Hong, P., Hong, S. W., Roh, J. J., & Park, K. (2012). Evolving benchmarking practices: A review for research perspectives. *Benchmarking: An International Journal*, *19*(4/5), 444–462. https://doi.org/10.1108/14635771211257945

Jonsson, P., & Lesshammar, M. (1999). Evaluation and improvement of manufacturing performance measurement systems: The role of oee. *International Journal of Operations & Production Management*, *19*(1), 55–78.

Leppävuori, I., Hinkel, J., Tonelli, M., Rutigliano, M., & Grant, B. (2026). *Paper & packaging report 2026: Adapting today, shaping tomorrow*. Bain & Company.

Li, X., Liu, G., & Hao, X. (2021). Research on improved OEE measurement method based on the multiproduct production system. *Applied Sciences*, *11*(2), 490. https://doi.org/10.3390/app11020490

Márquez Figueroa, L. J., García Alcaraz, J. L., Díaz Reza, J. R., & Quintana Alvarado, J. (2025). Gemba. In J. L. García Alcaraz, G. Cortés Robles, & A. Realyvásquez Vargas (Eds.), *Lean manufacturing in latin america: Concepts, methodologies and applications* (pp. 385–408). Springer Nature. https://doi.org/10.1007/978-3-031-70984-5_17

Ofori-Kuragu, J. K. (2025). Development of a conceptual benchmarking framework for the construction industry. *International Journal of Business Performance Management*, *26*(1), 83–103. https://doi.org/10.1504/IJBPM.2025.10059245

Olsson, H., & Sörensen, S. (2021). *Forskningsprocessen* (4th ed.). Liber.

Oral, M., Cinar, U., & Chabchoub, H. (1999). Linking industrial competitiveness and productivity at the firm level. *European Journal of Operational Research*, *118*(2), 271–277. https://doi.org/10.1016/S0377-2217(98)00354-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Gluud, C., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Porter, M. E. (1980). Industry structure and competitive strategy: Keys to profitability. *Financial Analysts Journal*, *36*(4), 30–41. https://doi.org/10.2469/faj.v36.n4.30

Puvanasvaran, P., Teoh, Y. S., & Tay, C. C. (2013). Consideration of demand rate in overall equipment effectiveness (oee) on equipment with constant process time. *Journal of Industrial Engineering and Management*, *6*(2), 507–524. https://doi.org/10.3926/jiem.537

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th ed.). Pearson Education Limited.

Slack, N., & Lewis, M. (2015). *Operations strategy* (4th ed.). Pearson.

Sreejesh, S., Mohapatra, S., & Anusree, M. R. (2014). *Business research methods: An applied orientation*. Springer. https://doi.org/10.1007/978-3-319-00539-3

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Trattner, A., Hvam, L., Forza, C., & Herbert-Hansen, Z. N. L. (2019). Product complexity and operational performance: A systematic literature review. *CIRP Journal of*

Manufacturing Science and Technology, *25*, 69–83. https://doi.org/10.1016/j.cirpj.2019.02.001

Ullah, M. R., Molla, S., Siddique, I. M., Siddique, A. A., & Abedin, M. M. (2023). Optimizing performance: A deep dive into overall equipment effectiveness (oee) for operational excellence. *Journal of Industrial Mechanics*, *8*(3). https://doi.org/10.46610/JoIM.2023.v08i03.004

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14)*. https://doi.org/10.1145/2601248.2601268

A

Appendix A: Method

This appendix contains the format for the questionnaire referenced in the Methodology Chapter.

Machine Performance Questionnaire

Industry benchmarking - product characteristics, KPI relationships and scenario reference speeds

B - Relationship matrix

For each product characteristic, rate the strength of its effect on each KPI (1–5) and indicate the direction of that effect.

Strength: 1 = very weak effect → 5 = very strong effect   Direction: ↓ = decreases the KPI | ↑ = increases the KPI | ~ = depends on other factors
** Run speed row: rate how run speed itself affects Waste Category 1 and Waste Category 2 (no entry needed under 'run speed' column).*

	Run speed		Waste Category 1		Waste Category 2	
Product characteristic	**Strength (1–5)**	**Direction**	**Strength (1–5)**	**Direction**	**Strength (1–5)**	**Direction**
Scale / direction key	*1 = very weak 2 = weak 3 = moderate 4 = strong 5 = very strong | ↓ = decreases ↑ = increases ~ = depends*					
Grammage	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~
Width	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~
Thickness	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~
Structure	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~
Run speed *	___	___	1 2 3 4 5	↓ ↑ ~	1 2 3 4 5	↓ ↑ ~

If you marked ~ (depends) for any cell, please briefly explain:

Which cell(s) and what it depends on:

C - Scenario interval definition

For each characteristic, enter the threshold values that separate the Simple, Medium and Complex scenarios. These will be used to define the product profiles for the benchmarking model.

Simple - A straightforward order that runs close to maximum machine speed.

Medium - An order that requires some operational adjustment, a moderate speed reduction. Still manageable within normal production flow but not a standard run.

Complex - An order that requires significant speed reduction, or combines multiple challenging characteristics simultaneously.

C1 - Grammage

Scenario	Lower boundary (gsm)	Upper boundary (gsm)
Simple	____________	____________
Medium	____________	____________
Complex	____________	____________

C2 - Width

Scenario	Lower boundary (% of machine max width)	Upper boundary (% of machine max width)
Simple	____________	____________
Medium	____________	____________
Complex	____________	____________

How does thickness determine scenario complexity in your view?

○ Thin = Simple / Medium; Thick = Complex	○ Thickness combines with other characteristics, does not determine tier alone

C4 - Structure type

Assign each structure type to a scenario tier based on its run speed characteristics. Mark one circle per row.

Structure type	Simple	Medium	Complex
Structure 1	○	○	○
Structure 2	○	○	○
Structure 3	○	○	○
Structure 4	○	○	○
Structure 5	○	○	○
Structure 6	○	○	○
Structure 8	○	○	○
Other: _______	○	○	○

Notes on any Complex structure types (what makes them particularly challenging):

Optional notes:

D - Industry reference speeds and order lengths

Please provide the best run speed range you consider achievable across the industry for each scenario for the year of 2025. Answer independently of any specific site, we are looking for an industry reference level.

D1 - Reference run speeds per scenario

Scenario	Best-in-class speed (m/min)	Typical speed (m/min)	Conditions / assumptions
Simple			
Medium			
Complex			

D2 - Waste Category 1 in % per scenario

Scenario	Best-in-class (%)	Typical (%)	Poor Conditions (%)
Simple			
Medium			
Complex			

D2 - Waste Category 2 % per scenario

Scenario	Best-in-class (%)	Typical (%)	Poor Conditions (%)
Simple			
Medium			
Complex			

E - Industry context

Q1. In your experience, which product scenario is most common across the industry, simple, medium or complex? Has this changed over the past 5–10 years?

Your answer:

Q3. In an industry context, what is the typical planned downtime rate (% of available production time) for the machine?

Your answer:

Q4. In an industry context, what is the typical planned downtime rate (% of available production time) for the machine?

Your answer:

Q5. For planned downtime, how many hours per year would you consider industry-standard for scheduled maintenance on for the machine?

Your answer:

Q6. In an industry context, what would you consider a typical unplanned downtime rate (% of available production time) for the machine?

Your answer:

B

Appendix B: Results

This appendix contains supplementary figures referenced in the Results Chapter.

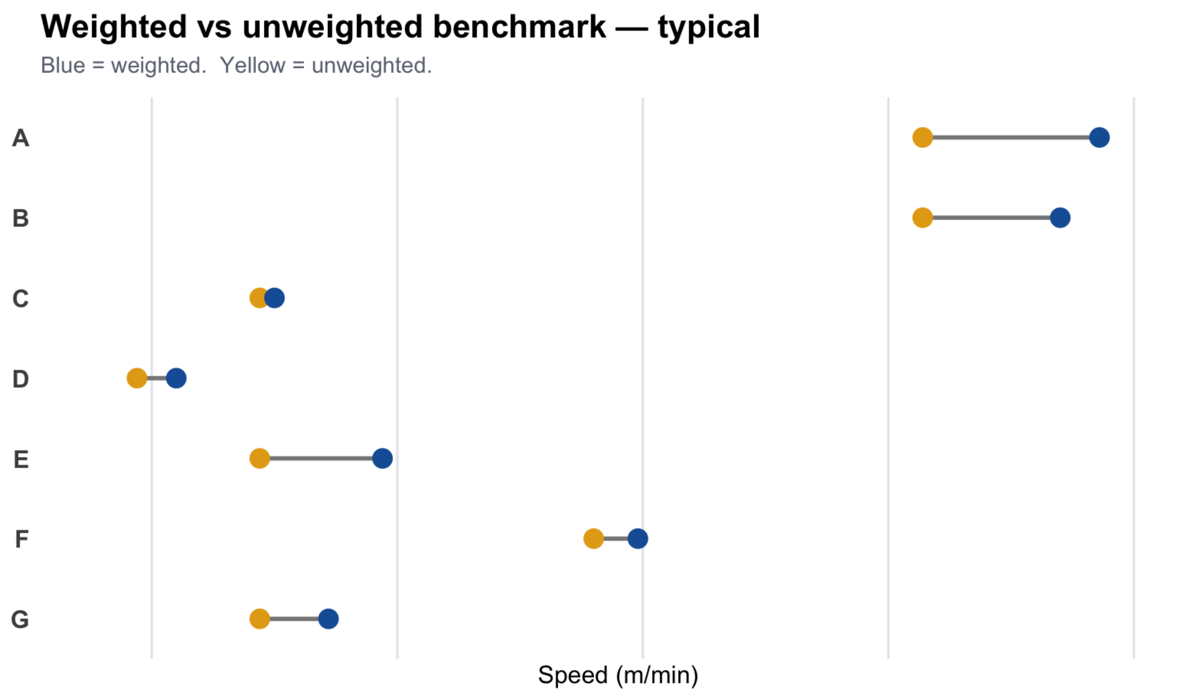


Figure B.1: Weighted vs unweighted typical industry benchmark speed per machine

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF INNOVATION AND R&D MANAGEMENT
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY