Tampere University

Sebastian Helala

DEVELOPMENT OF PERFORMANCE MEASUREMENT FOR THE PROCUREMENT OF PROJECT BUSINESS

Master of Science Thesis
Faculty of Management and Business
Examiner: Associate Professor Aki Jääskeläinen
Examiner: Professor Jussi Heikkilä
January 2022

ABSTRACT

Sebastian Helala: Development of performance measurement for the procurement of project business
Master of Science Thesis, 95 pages
Tampere University
Master's Degree Program in Industrial Engineering and Management
January 2022

Procurement is a key part of almost every manufacturing business. Successful supply chain management ensures that the required products or services arrive at the right time. To manage procurement effectively it requires implementing and following corporate strategies and procedures. Effective procurement performance measurement is needed to implement and monitor corporate strategies and targets. The thesis studies procurement performance measurement in a project business environment. The objective of the thesis is to develop a harmonized and coherent performance measurement system for one of the case company's business lines. Additionally, the objective is to align operational and strategical procurement targets and activities.

The research strategy for the thesis was the case study. With the case study approach, research can focus in-depth on the case and maintain a holistic and real-life viewpoint. Data for the research was gathered from qualitative and quantitative sources. Thus, the mixed method was chosen as a methodological choice. Qualitative data was gathered from semi-structured interviews and workshops. Quantitative data was gathered from the company's data cube. Induction was chosen as an approach for theory development.

At first, the thesis analyzed the current state of the case company's procurement performance measurement. The needs of the procurement stakeholders were gathered via interviews, workshops, and by examining previous performance measures. Draft dashboards were created based on the feedback from the interviews, corporate material, and literature. Created performance metrics were based on corporate strategies and objectives. The thesis uses a modified performance measurement system design process framework and implementation process framework in the implementation process. Visualization framework was used to determine the right visualization. These frameworks were used to create the dashboards. The draft dashboards were presented in several workshops, where stakeholders had an opportunity to propose changes and ask questions about the logic behind the measures.

The current status of the company's procurement performance measurement was overall in a good shape based on procurement coverage. Based on feedback from the interviews, the coverage of current measures was good. Only a few additional measures were requested. The major problems that occurred were, that a lot of manual work was required to obtain the measures. The second concern was the lack of harmonization between performance measurement systems between business units. The lack of harmonization restricted the cross-referencing and generated misunderstandings to the measuring logic between business units.

After the workshops, the draft dashboards were developed according to the requests given in the workshops. Several propositions were presented, which were evaluated according to the modified performance measurement system design framework. Finally, a total of three harmonized dashboards were created for the business unit management, category management, and team management. Transparency and harmonization were ensured by using the same dataset that was used for the organizational level. Literature frameworks conveyed the same message that was observed during the development process. The organization's strategy, objectives, and activities should be derived into the performance measurement. Thus, it can be used to monitor the strategy implementation, challenge the corporate strategy, and redirect focus on the critical areas.

Keywords: Performance measurement, procurement, supply chain, purchasing, implementation, development, project business, measurement, dashboard

TIIVISTELMÄ

Sebastian Helala: Hankinnan suorituskyvyn mittaamisen kehittäminen projektiliiketoiminnassa
Diplomityö, 95 sivua
Tampereen yliopisto
Tuotantotalouden diplomi-insinöörin tutkinto-ohjelma
Tammikuu 2022

Hankinta on keskeinen osa jokaista yritystä, joka toimii valmistavassa teollisuudessa. Onnistunut toimitusketjun hallinta varmistaa, että tarvittavat tuotteet ja palvelut saapuvat oikeaan aikaan. Hankintojen tehokas hallinta edellyttää yrityksen strategioiden ja menettelytapojen toteuttamista ja noudattamista. Yrityksen strategian ja tavoitteen toteutumisen seurantaan tarvitaan toimivaa hankinnan suorituskyvyn mittaamista. Diplomityössä tutkitaan hankinnan suorituskyvyn mittaamista projektiliiketoimintaympäristössä. Diplomityön ensimmäisenä tavoitteena on kehittää yhtenäinen ja harmoninen suorituskyvyn mittausjärjestelmä kohdeyrityksen liiketoimintalinjalle. Toinen tavoite on yhtenäistää operatiivisen ja strategisen hankinnan tavoitteet ja toiminnot.

Diplomityön tutkimusstrategiana oli tapaustutkimus. Tapaustutkimuslähestymistavan avulla tutkimus voi keskittyä tapaukseen syvällisesti ja silti säilyttää kokonaisvaltaisen ja tosielämän lähestymistavan. Tutkimusaineistot kerättiin laadullisista ja kvantitatiivisista lähteistä. Siksi sekamenetelmä valittiin metodologiaksi. Laadullista dataa kerättiin puolistrukturoiduista haastatteluista sekä työpajoista. Määrällinen data kerättiin yrityksen datakuutiosta. Opinnäytetyössä käytettiin pragmatismifilosofiaa. Induktiivinen päättely valittiin lähestymistavaksi teorian kehittämiseen.

Aluksi diplomityössä analysoitiin kohdeyrityksen hankinnan suorituskyvyn mittaamisen nykytilaa. Hankinnan sidosryhmien tarpeet kerättiin haastatteluiden, työpajojen ja kohde yrityksen aikaisempia suorituskyvyn mittareita analysoimalla. Haastatteluiden, työpajojen, yrityksen materiaalien ja kirjallisuuden pohjalta luotiin luonnosmittaristot. Luodut suorituskyvyn mittarit perustuivat yritykseen strategiaan ja objektiiveihin. Viitekehyksinä diplomityössä käytettiin muokattuja suorituskyvyn mittaamisen suunnitteluprosessin viitekehystä ja implementointi prosessin viitekehystä. Visuaalista viitekehitystä käytettiin visualisuuden tarkastelemiseen. Kyseisiä viitekehyksiä käytettiin mittaristojen tekemiseen sekä niiden kehittämiseen. Luonnosmittaristot esiteltiin useissa työpajoissa, joissa sidosryhmillä oli mahdollisuus ehdottaa muutoksia ja esittää kysymyksiä mittarien taustalla toimivasta logiikasta.

Yhtiön hankinnan suorituskyvyn mittaamisen nykytila oli kattavuuden osalta hyvässä kunnossa. Haastatteluiden perusteella nykyiset mittarit kattoivat hyvin hankinnan tarpeet. Haastateltavat pyysivät vain muutamia lisämittareita olemassa olevien lisäksi. Haastatteluiden ja työpajojen pohjalta suurimmat ongelmat olivat, että mittaustiedon kerääminen vaati paljon manuaalista työtä. Toinen ongelma oli suorituskykymittauksen harmonisoinnin puute liiketoimintayksiköiden välillä. Harmonisoinnin puute rajoitti yksiköiden vertailua ja aiheutti väärinkäsityksiä oikeasta mittauslogiikasta eri liiketoimintayksiköissä.

Työpajojen jälkeen luonnosmittaristoja kehitettiin työpajoissa esitettyjen toiveiden mukaisesti. Esitettyjä ehdotuksia arvioitiin muokatun suorituskyvyn mittaamisen suunnitteluprosessin viitekehyksen pohjalta. Lopulta yhteensä kolme harmonisoitua mittaristoa luotiin liiketoimintayksiköiden johtamiseen, kategorian hallintaan ja tiiminhallintaan. Läpinäkyvyys ja harmonisointi varmistettiin käyttämällä samaa tietolähdettä, jota käytettiin myös organisaatiotasolla. Kirjallisuuden viitekehykset olivat linjassa tutkimuksen käytännön havaintojen kanssa. Organisaation strategia, tavoitteet ja toiminnot tulee johtaa suorituksen mittaamiseen. Siten voidaan seurata strategian toteutusta, haastaa yrityksen strategiaa ja ohjata huomiota kriittisille alueille.

Avainsanat: Hankinta, hankintatoimi, mittaristo, mittaaminen, suorituskyvyn mittaaminen, johtaminen, projektiliiketoiminta, strateginen hankinta, operatiivinen hankinta

PREFACE

One of the key requirements for me has been that my master thesis should have some sort of useful impact. I want to thank Tuomo for offering the topic that fitted the requirement. I also want to thank you for all the support that he gave me during the process. The project taught me a lot about procurement, data analysis, and project handling. I got to know different people and their work from all levels of the company. The support and understanding that I got from everyone in the case company were really great, and I appreciate it a lot. I have learned so much and hope to learn even more. In the future, I hope to work with similar inspiring people.

Almost five years have gone by since I first started my studies at the university. It feels like yesterday. The years have been one of the most inspiring years I have had so far. So many memories throughout the studies and unforgettable events. I met many inspiring people and friends who pushed me forward. My friends really taught me how to aim high and more importantly how to work hard to reach those goals. I would not be in this situation without my friends. The university has taught me to be curious to learn new things and not to be afraid of hard topics. I value it highly. Special thanks to the Aki Jääskeläinen and Jussi Heikkilä for guiding me through the process. The advice and feedback that you gave me were really important. I want to thank you for your patience and interest in the topic.

Most importantly, I want to thank my family and friends who have supported me during the process. You have always wanted my best and were always ready to help. Thank you.

Tampere, 26 January 2022

Sebastian Helala

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

BL	Business Line divides the company's core competencies into different functions.
BU	Business Unit divides the business line into smaller functions, which are focused on core competencies.
CCC	Cost Competitive Country. Countries where products or services are more competitive to manufacture.
COGS	Cost of Goods Sold includes all the direct costs involved in the making of the commodity that is being sold.
ERP	Enterprise Resource Planning gathers information about the different functions of the company into one system. Functions can for example be finance, procurement, or production.
KPI	Key Performance Indicator shows how the corporation is performing in key success factors.
PI	Performance Indicator is a result of several teams working together. Shows what the teams are delivering but is not a key success factor.
KRI	Key Result Indicator shows the summary of how the organization is performing in key areas.
RI	Result Indicator is not so critical metric compared to the KRIs. It indicates how teams are combining to produce results. The time horizon is much wider, looking at the future months, weeks, and days
OTD	On-Time-Delivery measures if the good or commodity is delivered at the time which was agreed with the supplier.
YTD	Year-to-date shows the data from the start of the year to the current date. The cumulative sum is used to calculate the sum.
TCO	Total Cost of Ownership is the costs that occur on the whole lifetime of the commodity or service.

1. INTRODUCTION

In the introduction, the research background and motivation for the thesis are illustrated. The case company that requested the thesis is introduced. The thesis research objectives and questions are presented with the limitations and scope of the thesis. Finally, the structure of the thesis is explained.

1.1 Research background and motivation

Performance measurement can have a big impact on an organization and how successful it is. Selected measures have an impact on what aspects are followed and improved. Amongst the published works of the field, there are examples of when a set of performance measures have directed actions into the right areas. Weele (2018, pp. 161–162) mentions an organization that wanted to focus on improving four areas in the procurement: increase bundling by 60 %, increase global value sourcing by more than 5 %, reduce supply base by 20 % and increase e-auctions. The organization identified and built performance measures to monitor these areas. Strategies and action plans were put in place to improve the current values to the level that was required. These actions led to a 90 % increase in bundling, a 1000 % increase in e-auctions, global value sourcing increased more than 5 % and the count of suppliers decreased 20 %. The example illustrates the impact of performance measurement. By defining the crucial factors and measuring them, the organization could see how they are performing. By introducing targets to these crucial factors, illustrates a clear direction where the factors should be developed. These concrete targets drive action, provide feedback on performance, and motivate people to achieve these goals. In addition, measuring performance leads to better decision making. Actual results can be compared to planned results, and the variances can be analyzed to find the root cause and corrective action. (Weele, 2018, pp. 304–305)

Research in the field of performance measurement has made advances in the design and implementation process of the performance measurement system. However, certain areas still need more attention. These areas are the factors affecting the successful implementation of performance measurement, factors affecting the evolution of the measurement system, and problems in the maintenance of performance measurement systems (Bourne et al., 2000; Kennerley & Neely, 2003; A. D. Neely, 2002). The literature

has wide research on performance measurement frameworks but lacks the implementation phase of performance measurement (Bourne et al., 1999). There is a need for bridging the gap between literature and practice (Papakiriakopoulos & Pramatari, 2010). Therefore, this thesis is intended to bring more research into the implementation process and to bridge the gap between practice and literature.

From the case company's point of view, they had built a centralized "data cube" which would make measuring across the business lines (BLs) and units more coherent and transparent. Procurement was one of the first functions that was connected to this data cube. This opened a door for building performance measures for the procurement that would able accurate cross-comparison between business units (BUs). The performance measurement dashboards would provide more accurate data to help decision-making at the business unit level, category level, and operational procurement. Another driver for the case company to improve the performance measurement in procurement is that the use of performance measurement in procurement has been found to increase performance in procurement itself (Bello & Gilliland, 1997). Also, it has been argued that effective performance measurement sets the baseline for understanding procurement (I. J. Chen & Paulraj, 2004, p. 145). Actions and behavior can be influenced by the performance system, as illustrated in Weele's (2018, pp. 161–162) example. The second driver for the case company is that the company's performance measurement system would be modernized and partially automated, which would save valuable time for other activities.

1.2 Background of the case company

The case company for this thesis is a global supplier and developer of process technologies, automation, and services for the pulp, paper, and energy industries. The company employs nearly 14,000 people worldwide and is a leading manufacturer in those areas. The case company has identified a need for unified and reliable procurement measurement dashboards for its business units under the paper business line. Similar need rises for its strategic and operational procurement functions. Currently, the case company doesn't have unified and reliable measurements between business units, and the measurements are done mostly by hand. This weakens the reliability of the measurement and increases the workload for the managers. The current state of the performance measurement contradicts with the case company's current strategy of a unified and reliable performance measurement system.

The possibility for a new unified and reliable measurement arose from the case company's investments in digital solutions. The case company created a so-called "data

cube" where data is stored from all business units across the globe. This opened a possibility for creating unified measurement between business units. The business units that are on board in the unification process are TM, PM, SPR, and China Operations. These business units are illustrated in the Figure 1. Additionally, performance measurement is created for the category management (PAP Category Management in the Figure 1) and procurement team management.

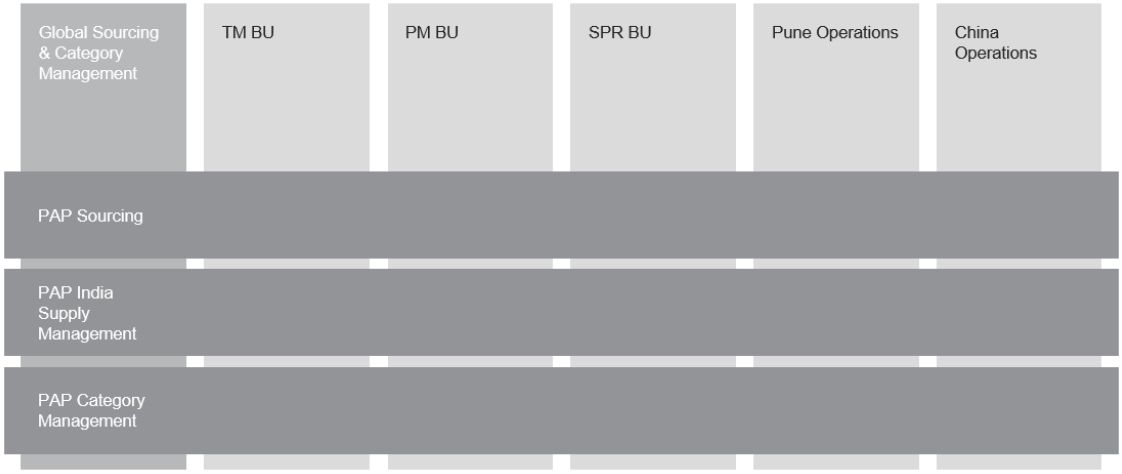

Figure 1. *Matrix procurement organization chart of the case function.*

In the case company and under the selected business line, responsibilities are divided similarly to matrix management organization structure. Matrix organization is described as a mixed organization form because the traditional hierarchy is overlayed by horizontal functions. The matrix organization structure is a mixture of functional organization and project organization structures. Projects are not divided as their own entities. Instead, project participants report to project and functional managers. Each team has its own responsibilities across the business lines. (Larson & Gobeli, 1987, p. 126) The organization functions under the business line are described as follows on the company site:

"The organization function is responsible for categories, where there are synergies in business line level, or that are critical to the success of business line".

Global sourcing in the Figure 1, makes sure that new geographical area opportunities are utilized optimally for benefiting the business line. The second responsibility is to govern that procurement strategy is implemented globally. Category management function's role in the Figure 1 is to form and implement long-term strategies aligned with global business needs to ensure a competitive and secure position in the supply market.

1.3 Research objectives and questions

The aim is to develop a performance measurement system for different stakeholders in the case company's procurement function. These different procurement stakeholders are the directors, category managers, procurement team managers, and buyers of the business units. The procurement organization function, which includes the stakeholders, is illustrated in the Figure 1. The goal of this thesis is to create dashboards for the business units and stakeholders inside the procurement organization function. The dashboard should include the performance indicators (PIs), result indicators (RIs), and specific metrics that were proposed by end-users. These measures should be aligned with the corporate strategies and objectives. The main objective of the thesis is:

"Create harmonized and coherent procurement performance measurement dashboards to reduce manual work and to align strategies and objectives of operative and strategic procurement".

Harmonization between different performance measurement dashboards allows comparison across the business units and ensures coherent measurement in the dashboards. This is one of the key areas that the dashboards must accomplish. Cross comparison allows to find inconsistencies and development areas from the procurement functions. For the team management performance measurement, one of the expected goals is to provide managers with tools to have better insight into the performance of the procurement team. Some of the business units had already similar performance measurement in place, but harmonized performance measurement across the procurement teams was missing.

The second key area is the reduction of manual work on operative and strategic procurement. By reducing unnecessary manual work on performance measurement, saved time can be used to create value somewhere else. Manual work creates possibilities for errors to occur which can lead to misunderstandings or wrong decision making. Before, a significant portion of the performance measurement was done manually. For example, directors were using a significant amount of time before the reporting periods to create performance measurement. Therefore, one of the expected outcomes is to reduce time on making performance measures. Gaining an understanding about the procurement performance measurement is also included in the objective. For gaining an understanding about the field, literature research and practical experience play a key role. The literature review is the foundation for the performance measurement development process. The implementing process and design process are researched from the field's literature.

Researched implementing process and design process are then modified to suit the case company.

The final objective for the thesis is to align the strategies and objectives of operative and strategic procurement. This is essential because then all the levels of the procurement are working towards the same goal. Additionally, by aligning the strategies and objectives on different levels of the procurement hierarchy, the end users are aware of what are the organization's strategies and objectives. This gives motivation for the operational procurement and aligns their work with the strategic procurement.

The main research question is focused on the development of a performance measurement system. The sub-research questions are focused on the current state of the performance measurement in the case company, why different procurement shareholders need performance measurement, what should be measured in project business, what performance measures procurement needs, and what are the development needs. The main research question for the thesis is:

Q1: How to develop a procurement performance measurement system in a company operating in a manufacturing project business?

The question has several sub-questions. They help to answer the main research question by dividing the topic into smaller sections. The first sub-question is:

SQ1: What is the current situation of the procurement performance measurement in the case company?

This sub-question is intended to find the current state of the procurement performance measurement in the case company. What are the challenges that the case company faces in the area? What are the current developments? The question lays the foundation for the following sub-questions. By understanding the current status, the next phase is to understand why performance measurement is needed and the needs of different stakeholders. Those findings can be compared to the current status in the case company. Thus, the following sub-question is focused on the reasons why performance measurement is needed in procurement:

SQ2: Why the internal procurement stakeholders need performance measurement?

The intend of this question is to find out the reasons why performance measurement is needed in procurement. The internal stakeholders are the before-mentioned business line directors, category managers, team managers, and buyers. The final sub-question focuses on the case company's procurement and its needs:

SQ3: What kind of performance measures does the case company need to support strategic and operational procurement? How they should be developed?

The focus is around the case company and its needs for supporting strategic and operational procurement. What are the aspects that a good measure has that support the strategic and operational procurement? The intent is to answer what are the common characteristics that a good measure has. What are the tradeoffs that had to be made?

1.4 Scope and delimitation

The scope of the thesis is about direct procurement. The case company has an interest in measuring direct procurement after the creation of the data cube. Thus, indirect procurement is excluded from the scope of the thesis. Due to accurate data on the India business unit, because of different enterprise resource planning (ERP) systems, it was excluded from the scope of the thesis. Therefore, in the Figure 1, the Pune Operations (India) is not mentioned.

The strategic, tactical, and operational procurement are defined in the thesis as follows. Top management, logistic management, and procurement management are responsible for the strategic level. The decision on the strategic level includes investment decisions, outsourcing decisions, and developing procurement frameworks. The time horizon for the strategic level is long. The tactical level works as a bridge between strategic and operational levels. Category managers, team managers, and senior buyers operate at the tactical level. Their responsibilities are supplier audits, supplier agreements, and supplier selection. The timeframe for the tactical procurement is mid-range. The operational level consists of buyers and team managers. They release the purchase orders, follow on-time delivery (OTD), and troubleshoot problems. The operational level ensures that the commodities and services are delivered on time and with the required quality. They operate in a short timeframe. (Schmidt & Wilhelm, 2000) The definition for performance is referred to as the achievement of a result. In addition to achieving results, performance is defined as a process of performing something. What is being done to achieve the result? Finally, performance is defined as a perceived observation of how the process was performed. In the literature supply chain, procurement, and purchasing are used sometimes as synonyms. In this thesis mainly the term procurement is used.

In the performance measurement design phase, a modified version of the Neely et al. (2000) framework for the performance measurement design process is used. The framework was chosen because it had practical aspects and was comprehensive. The framework assessed the common findings of the performance measurement design literature.

Additional feature was added to the framework from Kaplan's and Norton's (1992) balanced scorecard. The added feature gives a more balanced approach to the performance measurement design. The actual steps for developing the performance measurement are derived from the Parmenter's (2020) performance measurement implementation framework. It is modified to suit the case company and the situation. For the design of the measures, Abela's (2008) visualization framework is used. The framework is used to define appropriate visualization for the required measures.

This thesis takes into consideration the internal and external procurement performance perspectives. Not all aspects are considered in the procurement performance, due to limited resources. The categories of procurement performance are illustrated in the Figure 6. From those categories, procurement effectiveness and efficiency are included partially as part of procurement performance. The procurement material price, procurement logistics, supplier performance are included in the procurement effectiveness category. Supplier performance is only limited to current suppliers and to the already present measures. Due to the limited time and resources, supplier performance monitoring and supplier management are out of the scope of this thesis. Product quality is not included. Quality measurements are only mentioned briefly in the thesis. This is due because of the lack of accurate quality data. The data on quality was not implemented into the data cube that the case company has developed. On the efficiency side of the Figure 6, the emphasis is on better management of the procurement function through the performance measurement dashboards.

In this thesis, the implementation frameworks are derived from theoretical backgrounds and the implementation and analysis of the performance measurement system are done based on the frameworks. Thus, the scope does not cover the whole procurement measurement. The covered procurement measurement is brought up in the theoretical background chapter.

Data validation is out of the scope of this thesis. The data cube created by the case company has data validation already in place. Thus, data validation is not done in the thesis, and it is left to the key users of the different business units. Some data validation is done between and after the workshops. Due to a limited timeline, the maintenance of the performance measurement is not researched in this thesis. Channels are created so that the end-users can inform the developing team about updates and possible problems. Regular maintenance and update meetings are planned to keep the performance measurement dashboards up to date. Only the business unit dashboard is assessed more closely due to the limited space of the master thesis. Totally of three dashboards are

created: business unit, category management, and team management. The development processes are identical between the dashboards, but only the business unit dashboard is illustrated in the thesis. The dashboards illustrated in the thesis are created in a demo environment, no real values are used.

1.5 Structure of the thesis

The thesis starts with an introduction of the case company. Research questions and objectives are presented, and the scope is defined. After the introduction, the focus switch towards the theoretical background. In the theoretical background, the procurement function and strategies are presented, performance measurement is defined with the indicators, and procurement performance is illustrated. Additionally, the need for performance measurement is researched. Implementation of performance measurement system is illustrated. Finally, dashboards are illustrated with the visualization of measures. The theoretical background includes the research of the mentioned frameworks for the performance measurement implementation process and design process. After theoretical background, the research methodology is defined. This defines the methods, strategy, and timeline used in the thesis. Data gathering methods are introduced. Data is gathered from interviews, workshops, personal meetings, and the company's data warehouse. Additional data is gathered from company materials. Interviews and workshops are reported, and participants are presented. Interview questions are also illustrated.

The results and analysis chapter illustrates the results from the development process of the performance measurement. The current status of the case company's procurement performance measurement is presented. Findings from the interviews and workshops are presented. Feedback's effect on the development process is presented. Frameworks from the theoretical background are used in the implementation process. Finally, the actual development process is presented with its results (dashboards). Created dashboards are illustrated and analyzed.

In the final chapter, the findings from previous chapters are compared to the theoretical research. The implemented development process is compared to the implementation process framework and their key findings are analyzed. In addition, the performance measurement design process framework is compared to the findings in the development process. Visualization framework is reviewed and compared to created dashboards. After the key findings, managerial implications are presented. Next, criticism and limitations of the thesis are illustrated, and future development paths are presented.

2. THEORETICAL BACKGROUND

In this chapter procurement as a process, procurement strategy, performance measurement, and need for it are researched from the literature. Performance measurement is assessed at the general level and at the procurement level. Additionally, performance and result indicators are researched, pitfalls are illustrated, the implementation process of performance measurement is illustrated. The frameworks that are used in the thesis are illustrated. The last sub-chapter asses' procurement dashboards.

2.1 Procurement

Procurement is a key part of almost every manufacturing business. Successful supply chain management ensures that the required products or services arrive at the right time. This ensures that production can run smoothly. To manage procurement effectively it requires implementing and following corporate strategies and procedures.

Traditionally the procurement function is divided into direct and indirect procurement. In this thesis, only direct procurement is assessed. Direct procurement means materials, products, and services that are used directly in the primary activities of the business. An example of direct procurement is the purchase of bearings that are assembled into the product. On the contrary, indirect procurement is the materials, products, and services which are directly used internally to develop the company. These are for example IT systems. Direct procurement is in the scope of this thesis and indirect procurement is ignored. (Weele, 2018, p. 6)

In the literature supply chain, procurement, and purchasing are used sometimes as synonyms. To clarify the terminology around the procurement Weele (2018, p. 8) formed a linear procurement process model. As mentioned, this model illustrates the terminology but also the structure of activities that comprise the purchasing and supply management. The model is illustrated in the Figure 2 below.

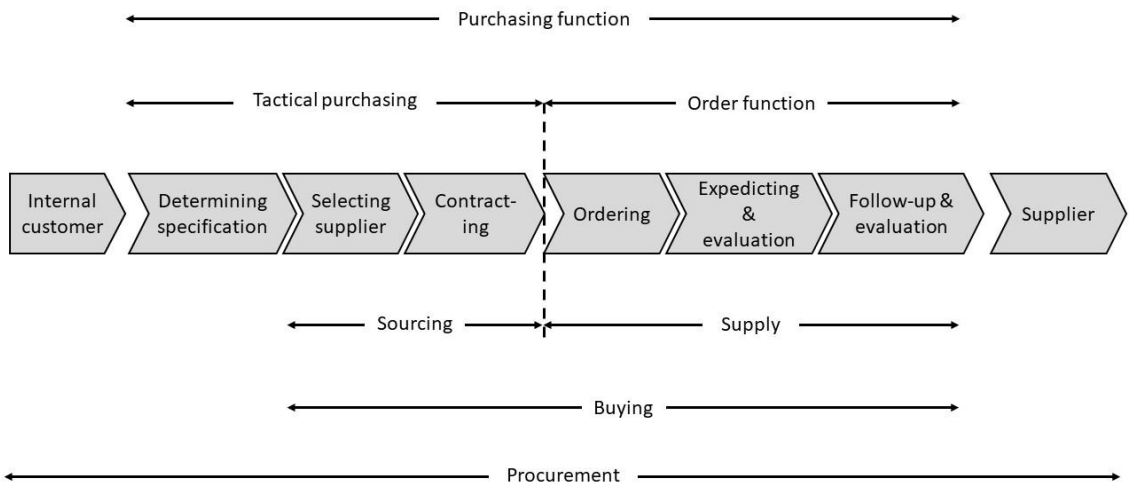

Figure 2. *Linear procurement process model.* (Weele, 2018, p. 8)

The linear procurement process model, illustrated in the Figure 2 above, starts with a make-or-buy decision in the company. When a decision of buying a commodity or service is done, the determination of variables starts. These variables are quality and quantity. After the determination process, the best suitable supplier is selected and procedures for selecting suppliers are developed. Then a decision on the type of contract is made and the contract is signed with a supplier. Purchase orders are put in place and follow-up of the order is conducted. Some of the followed attributes are on-time delivery and lead times. The supplier's invoice is checked against the purchase order. Nowadays many of these procedures are automated. As illustrated in the Figure 2, buying consists of sourcing and supply chain. As illustrated in the Figure 2, procurement is responsible for supporting the company's primary activities most optimally and securely possible. This includes delivering goods and services at the right moment and with the right quality, examining the right suppliers for certain commodities, managing supplier relationships, negotiating agreements with the suppliers, make-or-buy decisions, and monitoring and controlling orders. (Weele, 2018, p. 7)

The extended version of Weele's (2018) procurement process model illustrated in the Figure 3 is more comprehensive but fundamentally is similar to the linear model. The most notable difference is the follow-up after the contract. Supplier invoices are compared with the contractual obligations to ensure that the company pays on the performance that it has agreed. (Weele, 2018, p. 33) This is illustrated in the center of the model, which was modified to include supplier development, relationship management, and evaluation by Raaij (2016). The center illustrates the constant supplier evaluation,

development, and management. The stakeholder management assesses the interest and influence of stakeholders purchasing decision-making outputs. Risk management, illustrated in the Figure 3, includes the actions towards reducing unwanted risk.

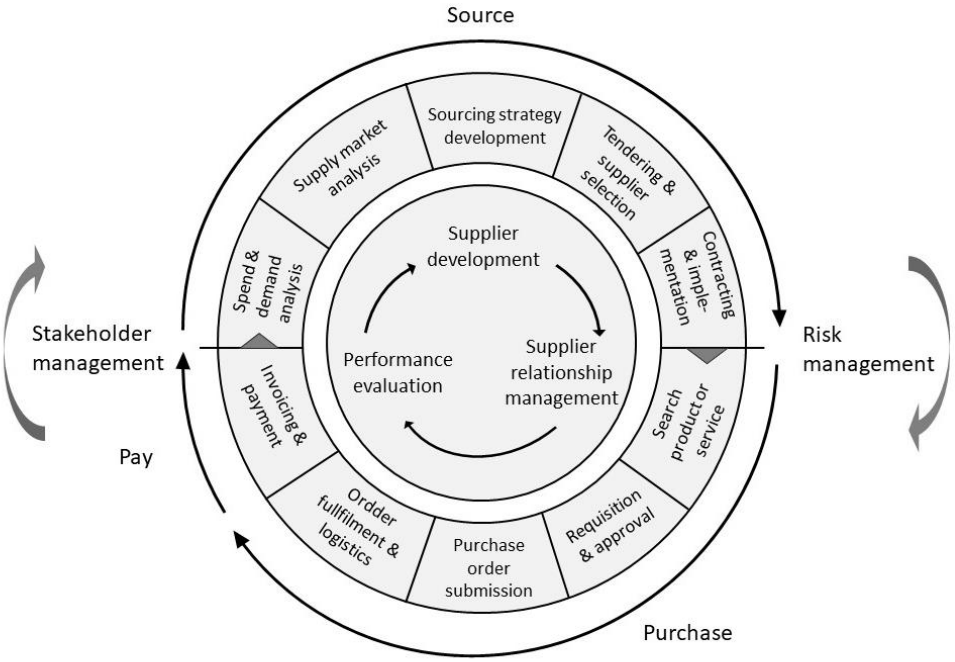

Figure 3. *Extended purchasing process* (adapted from van Raaij, 2016)

The model illustrated in the Figure 3 lacks supplier exit aspects (Bäckstrand et al., 2019). Even though there is a lack of supplier exit, the extended and linear model gives a comprehensive understanding of procurement's actions.

2.1.1 Procurement in project business/turnkey

The case company focuses on providing turnkey solutions to the customers. This means that the company provides different products combined with a service to add value to the customers. In turnkey projects, the company takes more risk for the success of the project. This opens the door for innovations leading to developing new ways of using products together, thus adding value. (Brady et al., 2005) The highly complex product environment that the case company operates in has a direct effect on procurement. The company has a strong focus on innovations and cost reduction. The procurement will be working with detailed material budgets and cost reduction projects. This shifts procurement to constantly search for competitive suppliers. (Weele, 2018, p. 65)

Taking into consideration that turnkey projects are usually more complex than non-turnkey projects, it affects the procurement also. Tendering-based purchasing is preferred in the turnkey project environment. The primary reason for this is that the suppliers compete

against each other, leading to lower prices. In complex turnkey projects, the comparison between different suppliers is difficult, preferring a more flexible tendering mode. The downside of competitive bidding is that it requires a lot of resources and leads to high transaction costs for the customer. (Ahola et al., 2008)

A different approach is to form a close relationship between buyer and supplier. In this context, the participants consider each other more as a partner The close relationship leads to several benefits. Ahola et al. (2008) and Giunipero (2000) mention the following:

- High cost of competitive bidding is eliminated.
- Fever suppliers lead to a reduction in supplier management costs.
- Mutual beneficial agreements are possible.
- Risk is shared between parties which might lead to innovations.

Closer relationships have also their drawbacks. It might lead to favoring a supplier over another, hindering innovation (Ahola et al., 2008). The supplier might use the opportunity of reduced competition and raise prices.

2.1.2 Procurement strategy

Increasing global competition has put increasing performance improvement demand on businesses. The need for continuous improvement has driven demand for global corporate strategy. This in turn has led to demand for global sourcing strategies, because of procurement's high spending. (Petersen et al., 2000) On average, a manufacturing firm spends 55 % of its revenue on purchasing (Tully, 1995). Effective procurement helps in the upturn economy to increase profit margins and in the downturn, reduce the profit reduction, while keeping quality at an acceptable level (Dimitri, 2013). Casson (2013) argues that fundamental economic factors have ultimately an impact on procurement parameters, defining its shape, size, and characteristics. Additional drivers are argued to be regulatory frameworks, technology breakthroughs, sustainability, political factors, and strategic choices (Ketchen & Giunipero, 2004; Louviere et al., 2007; Türkay et al., 2016; Woody, 2012). To manage these drivers and factors, a global sourcing strategy has been seen as critical to a company's success (Petersen et al., 2000).

The procurement of global organizations makes decisions at strategic, tactical, and operative levels. Top management, logistic management, and procurement management are responsible for the strategic level. The decision on the strategic level includes investment decisions, outsourcing decisions, and developing procurement frameworks. The time horizon for the strategic level is long. The tactical level works as a bridge between

strategic and operational levels. Category managers, team managers, and senior buyers operate at the tactical level. Their responsibilities are supplier audits, supplier agreements, and supplier selection. One limiting factor towards the tactical level is the network availability based on the strategic-level decisions. The timeframe for the tactical procurement is mid-range, from 6–24 months. On the contrary, the operational level operates under the decision created by the tactical level. The operational level consists of buyers and team managers. They release the purchase orders, follow on-time delivery, supplier evaluation based on performance, and troubleshoot. The operational level ensures that the commodities and services are delivered on time and with the required quality. The operative procurement operates in a short timeframe. The uncertainty of the decision grows from bottom to top levels. (Schmidt & Wilhelm, 2000)

2.1.3 Purchasing hierarchy

In the literature, there has been a lack of a clear classification of hierarchy between procurement strategies. The term “purchasing/procurement strategy” is used vaguely without a proper understanding of the principles behind it. No single strategy can manage different categories and suppliers. Therefore, strategy has to be implemented in hierarchies and plans to the different levels of procurement so that the strategy is executable (Nollet et al., 2005). Thus, Hesping and Schiele (2015) introduce a 5-level hierarchy to develop purchasing strategy. The 5-level hierarchy is illustrated in the Figure 4 below.

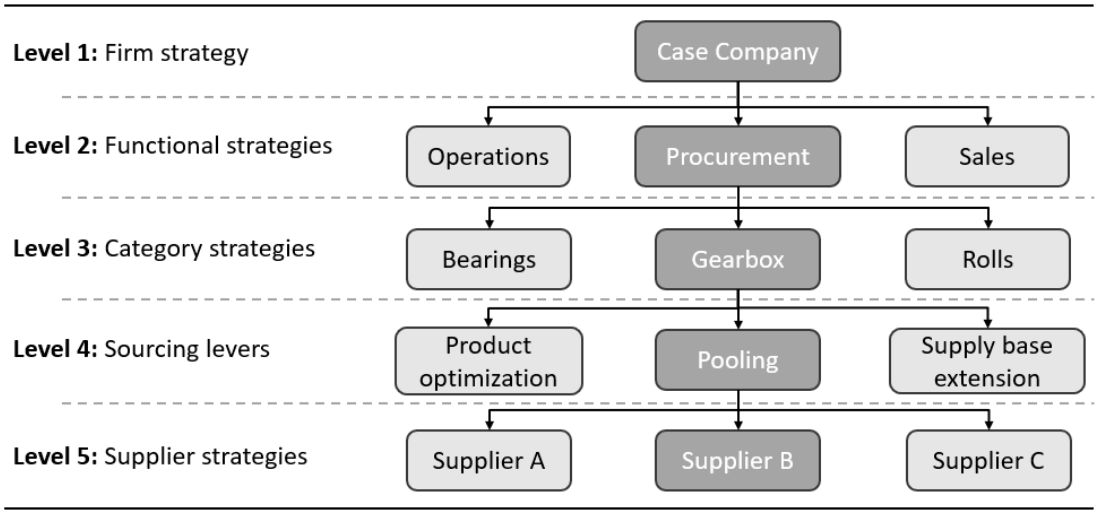

Figure 4. *5-level purchasing strategy development* (adapted from Hesping and Schiele, 2015).

The first level introduced in the hierarchy model is the firm strategy level. It coordinates the activities in the company and acts as a guideline between functions. These functions can be for example procurement, operations or sales. These functions and their strategies are illustrated in the next level. (Hesping & Schiele, 2015)

Functional strategy is an inner strategy that is guided by the firm strategy. General guidelines from the path to which the company is wanted to be steered are given from the firm strategy level. Then the function strategy is formed to accomplish the firm strategy at the functional level. (Hesping & Schiele, 2015)

Firm strategy can for example be that the company wants the on-time delivery to be close to 100%. At the functional strategy level, this means that the operations function might have a functional strategy that increases safety stock. This in turn urges the procurement function to increase relationships and integration with suppliers to meet the firm strategy of high OTD (Forslund & Jonsson, 2010).

The third level, as illustrated in the Figure 4, is the category strategy level. Category strategy is formed to differentiate the strategy between different supply markets. The strategy takes into account the functional strategy. Strategies are formed for homogenous products and services. Hesping and Schiele (2015) point out that the homogenous products and services are not to be mixed with a single product or raw material but represent groups of goods procured from similar suppliers overlapping each other. This has not always been the case and has been overlooked in the literature. Kraljic (1983) uses the “Kraljic matrix” where materials, under the same group, are placed in the four fields to analyze sourcing category strategies for different markets. Kraljic (1983) uses an example of oil products for different vehicles: jet fuel, trucking fuel, and shipping fuel. Hesping and Schiele argue that all these different uses from the same source should be under one category not four as Kraljic presents (Hesping & Schiele, 2015).

From the category strategy level, the next level is the sourcing levers. Hesping and Schiele (2015) illustrate this level as building blocks for the upper levels. Strategy levers help to transition the category level strategy into developing activities when choosing suppliers and monitoring supplier performance. Levers can focus on monetary measures, innovations, or adversarial. An example of a good sourcing lever for cost savings would be the project improvement lever. Other sourcing levers could be “supply base extension”-lever or pooling lever. Levers are best suited to describe basic elements, principles, and actions to make the best possible purchase from different suppliers (Schuh & Bremicker, 2005). One could assume that the levers might be useful in operational procurement dashboards, for example, team dashboards or buyer dashboards.

The final stage of the hierarchy is the supplier strategy level. Although all the suppliers are under the same category, supplier strategies between suppliers differ from each other. The category and supplier strategies are affected by the type of purchase and its requirements (lead time, order amount, relationship). (Stolle, 2008) The supplier strategy

guides the category manager on how to manage suppliers in mind of category strategy (Weele, 2018, p. 215). The “supply base extension”-lever could differentiate strategies between suppliers. For example, if one supplier’s volume is reduced, this freed volume can now be redirected to cost competitive country (CCC) supplier. For quality reasons, some volume could also be transferred to a more reliable supplier. These actions widen the supply base and reduce costs, which in procurement is usually one incentive. As seen from the Figure 4, the sourcing lever might affect several suppliers and thus it is different from the supplier strategy. (Hesping & Schiele, 2015) In the literature, it is noted that when the purchasing and supply management practices are in line with the corporate strategy, they can outperform the counterpart that focuses only on their objectives (David et al., 1999).

2.2 Category management

Category management is a dynamic process of data gathering, market analyzing, and market data reviewing to generate and implement spend strategies to benefit the organization in the long run (Cordell, 2018). Category management derives the strategic goals of the company by managing the supplier base with category-specific sourcing strategies (Weele, 2014, pp. 193–200). According to Heikkilä et al. (2018), the objective of the category management is to cross-functional integration, pool resources together that are sensitive to the volume, and support other business functions. Category management can be used as an organizational design element and in more complex environments as an integrative force to manage complex purchasing tasks. The role of category management varies according to the organization’s size. Cordell (2018) mentions six cornerstones towards successful category management:

- Customer focus. Every objective or target must be customer-led. The customer can be internal or external.
- Category management is about change.
- Process approach. Activities are divided into a smaller processes to be more systematic.
- Cross-functional approach. Category management works with a wide range of functions and stakeholders.
- Data and fact driven. Decisions are based on data and facts. Removes subjectivity.

- Continuous improvement. An infinite number of iterations to seek and implement improvements.

Category management has two premises: portfolio approach and purchasing synergies. These introduce a strategic element to category management. The before-mentioned portfolio approach means that items are analyzed and categorized. Individual purchasing strategies are created for each group. (Kraljic, 1983) This approach is also approved by other literature. Weele (2014) mentions in his study that the more homogeneous the item groups are, the more benefits can be found from the category approach. The latter premise of purchasing synergies means that when multiple business units combine their purchasing power to gain competitive advantage through cost-effectiveness. Synergies are achieved by exploiting interrelationships, knowledge sharing, resource sharing, coordinating strategies, and pooling negotiation power (Faes et al., 2000). In literature, synergies are further divided into three categories: economies of scale, economies of information and learning, and economies of the process (Trautmann et al., 2009). The first category, economies of scale increase purchasing power by pooling volume items and standardization of categories. The second category, economies of information and learning refers to sharing of information and knowledge across the different functions and locations. This means, for example, new technologies, best practices, and experiences. The third category, economic process means that the common way of doing things is established across the company. (Faes et al., 2000; Rozemeijer, 2000; Trautmann et al., 2009)

The second strategic process of category management includes linking procurement to the objectives and targets of the organization. Category management helps to convey the strategy to the lower levels of procurement and therefore it can be considered as a strategy process. The strategy process of category management and the whole procurement function is illustrated in the Figure 4. The category strategy adapts the procurement strategy to the different markets. In the literature, it has been pointed out several times that categories need specific category strategies and supply market strategies from the company strategies (Essig, 2011; Kraljic, 1983). Category strategies should follow the guidelines of coherent and integrity of functional strategy and make certain that the purchasing strategy and category strategy support the firm and business strategic objectives (Nollet et al., 2005).

In practice category manager’s duties varies between organizations, but similarities can be found. The basic duty is to manage the category that the category manager is responsible for. In Heikkilä’s and Kaipiala’s (2018) research, the category manager is involved in development projects within cross-functional teams, responsible to create category

strategy, developing the category, communicating within the category and to the management, and doing contracts with the suppliers. Category managers depend on good data and information from the supplier and the company. Dependency on good data is illustrated in the literature. Weele (2018, pp. 218–219) mentions that the category manager is responsible for monitoring supplier performance based on agreed key performance indicators (KPIs) with the supplier. In frequent supplier review meetings, KPIs are compared to agreed ones and the issues are analyzed. Thus, errors in the data are not tolerable. Another example of dependency on good data is the category manager's role in supervising contract compliance. Contract compliance monitoring is one of the category manager's responsibilities. The category manager makes sure that the company follows the agreed agreement. Example of contract compliance is to monitor payment terms.

2.3 Performance measurement

Performance measurement measures certain functions or entities based on objectives and targets agreed in advance. With the performance measurement system, a bigger picture from the function or entity is formed. This bigger picture is intended to help in decision-making by highlighting problematic areas. According to Komatina et al. (2019), the fundamental objective of the performance measurement is to indicate areas that do not reach the target figures. Kerzner (2013, pp. 117–118) refers to performance measurement as an early warning sign that if left untended may lead to unwanted situations. Thus, it increases visibility in the organization. Performance measurement can also indicate areas that are succeeding their targets and thus be an indicator for bonuses. Bonuses and performance measurement generally tend to motivate into actions and help people to see the progress and trends. They improve consistency and give early warning to problems that need focus (Parmenter, 2020). Performance measurement itself consists of several metrics and indicators. It can be assembled into a dashboard, scorecard, or report. In the thesis, the focus is on dashboards. Measures can be divided into financial, non-financial, internal, external, efficiency, and effectiveness measures (Kaplan & Norton, 1992). These metrics are chosen based on corporate strategy. A good performance measurement system conveys the corporate strategy to the function that it measures (A. D. Neely, 2002). Therefore, it is critical to consider this when designing a performance measurement system.

In the scientific literature, there are several approaches for designing performance measurement. Different frameworks are created to ensure a balanced view of performance.

One of the most famous frameworks for performance measurement has been the balanced scorecard by Kaplan and Norton (1992). The balanced scorecard links the measurement to the organization's strategy. The scorecard measures organization in a more balanced manner. It takes into account financial, non-financial, internal, external, efficiency, and effectiveness measures. Another approach for designing performance measurement in the literature is to set certain criteria that good performance measurement has (A. Neely et al., 1997). With this approach, the performance measurement can be matched with organizational content. According to literature consensus, good performance measurement should include two factors. First performance measurement is unique to the organization and thus, should be derived from organization strategy, vision, mission, and goals (A. Neely et al., 2000; van Looy & Shafagatova, 2016). Secondly, financial, and non-financial measures should be included (van Looy & Shafagatova, 2016). In the other publications, it has been found out that large and smaller companies tend to use both financial and non-financial measures, although large ones prefer financial measures (Malina & Selto, 2004a; Richard et al., 2009). Neely (2002) found similar characteristics for the performance measurement system in the literature research. Additionally, an organization needs to have comprehensive and multi-dimensional performance measurement. It should similarly be integrated across the organization's functions and hierarchies. This would allow cross comparison. Finally, there is a need to measure results and in drivers behind the results for analyzing past performance and forecast (Brignall & Ballantine, 1996). There has been a debate between the written works about the use of objective and subjective measures. Objectivity is preferred by the authors, but it is argued that with subjective measures the end users are involved more closely in the development of performance measurement (van Looy & Shafagatova, 2016).

Based on the literature analysis, the framework created by Neely et al. (2000) is used and modified to suit the thesis' procurement performance measurement. As illustrated above, in the literature there are different frameworks and each of them have their own nuances. Neely's framework was chosen because it had practical requirements and it was comprehensive. The framework is illustrated in the Table 1.

Table 1. *The framework for the performance measurement system design process* (adapted from Neely et al., 2000).

Desirable characteristics of a performance measurement system design process	**Desirable characteristics of the output of the process**
Performance measures should be derived from the organization's strategies and objectives	Performance measures should enable benchmarking
The purpose of each performance measure must be made explicit	Ratio based performance measures are preferable to absolute numbers
Stakeholders (end users, managers, and customers) should be involved in the development process	Performance measures should be directly under the control of the evaluated organizational unit
Data collection and the logic of calculating the performance must be clear	Objective performance measures are preferable to subjective ones
The selected performance measures should take account the organization	Non-financial measures should be adopted
The process should be easy to revisit – measures should be changed as the surrounding changes	Performance measures should be easy to use and simple
	Performance measures should provide fast feedback
	Continuous improvement should be stimulated rather than just monitoring
	Performance measures should be balanced between internal, external, efficiency, and effectiveness

The framework includes all the major criteria for the performance measurement design process. As illustrated in the Table 1, the framework takes the organization's strategies and objectives into an account. The framework tries to limit the number of performance measures by explicitly explaining the purpose of each measure. It emphasizes the calculation logic behind the measures and assesses the maintenance of the performance measurement system. Expected outputs of the design process are objective, easy to use, easy to read, and benchmarkable measures. Non-financial measures should be included in the process. Additionally, the measured function should have control over what measures are included in the performance measurement dashboard. (A. Neely et

al., 2000) The original framework did not include the Kaplan's and Norton's (1992) balanced approach of including internal, external, efficiency, and effectiveness measures. In the balanced scorecard framework, the balanced approach gives managers a fast and comprehensive view of the situation in the company. If the balanced view is preferred performance has to be measured on all aspects. In the literature, performance measures are often categorized under effectiveness and efficiency (A. Neely et al., 2005). Similar categorization is done by Weele (2018, p. 307) with the procurement performance, which is illustrated in the Figure 6. In the Weele's approach, procurement performance is derived into procurement effectiveness and efficiency. Another approach in the published works has been to categorize performance measures between the external and internal measures (Jääskeläinen, 2018; Pohl & Förstl, 2011). Therefore, effectiveness, efficiency, external, and internal measures have been seen as an important features in the performance measurement design when the balanced view is preferred.

The common perception is that all the performance measures are KPI's. That is not the case. Parmenter (2020) brings up a couple of measures that are usually mistaken as a KPI. One for example is revenue per employee. In the context of procurement, this could be the Purchase Order Amount per supplier. This kind of measurement is a result indicator. It reflects only results from different actions, for example, discounts to the supplier or seasonal buying. In Parmenter's (2020) book, performance measurement is divided into two main categories. With these main categories, there are two sub-categories for the measurement. The categories are illustrated in the Figure 5.

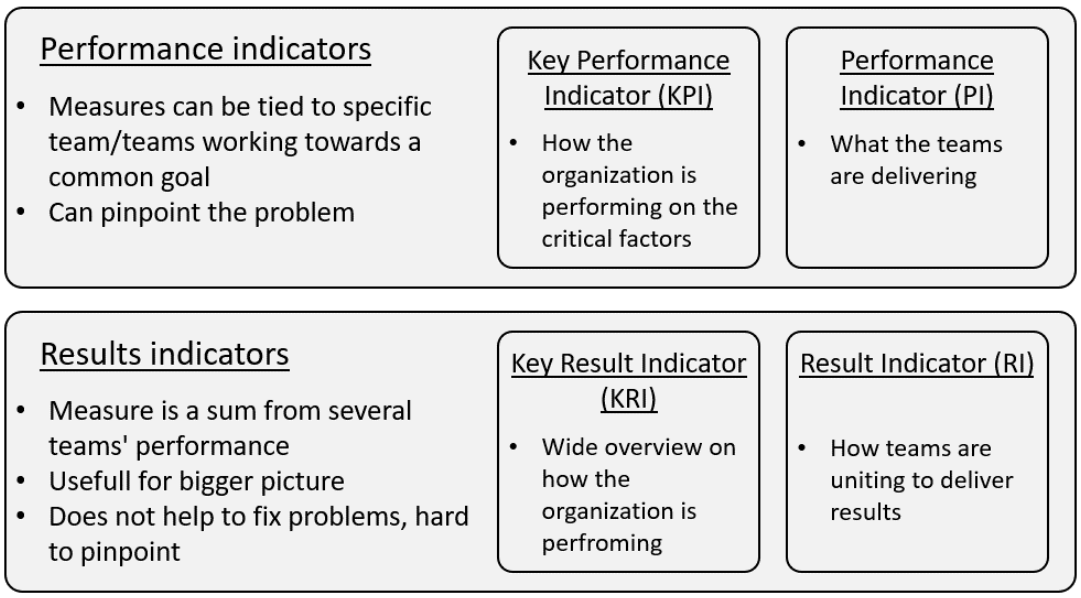

Figure 5. *Different types of performance measures* (adapted from Parmenter, 2020).

As illustrated in the Figure 5 the different groups have different meanings, which are sometimes mixed in the corporate world. A performance indicator is a result of one or more teams working closely together. Its result can be pinpointed into those teams. They are responsible for its success. Corrective measures can be identified, and corrective measures can be directed straight to the team. Result indicator consists of several groups' results towards a target. Its result cannot be pinpointed to one team. Similarly, no single group is responsible for its success. Result indicator gives wide inside into corporate's performance on a certain area. Direct corrective actions cannot be made solely based on the result indicators. All metrics that measure financial results are result indicators. The wording "key" in front of the indicators, means that the indicator measures key success factors for the company. (Parmenter, 2020) An example of a key performance indicator in the case company is the measurement of OTD. Key result indicator is for example procurement savings metric or CCC % (Cost Competitive County) metric. One team cannot solely change the result of the CCC metric. It is a group effort. Ultimately the CEO or the board of directors are responsible for this metric. This is a good indication when a metric is a result indicator. The responsible person is the CEO or the board.

2.4 Procurement performance measurement

As mentioned, the performance measurement system has been found to increase the performance of the function that it measures. As procurement accounts for the major share of the manufacturing company's expenditure, increasing its performance can lead to major savings (Tully, 1995). Bello et al. (1997) found out that the performance measurement system increased the performance of the procurement function when the system was implemented in the function. As a major part of manufacturing companies' expenses comes from procurement this increases the importance of the procurement performance system (Tully, 1995).

Usually purchasing performance consists of measuring costs, time, and quality (Luzzini et al., 2014). Traditionally cost has been the dominant measure to indicate procurement performance. This approach leaves flexibility, sustainability, and innovation out of the equation, which nowadays are important factors indicating procurement performance. Additionally, only internal factors are focused on when procurement performance is derived (Caniato et al., 2014). One could argue that procurement performance is also indicated by how well the corporate strategy is aligned to the procurement strategy and category strategies. Performance measurement has an important role in aligning these different strategies together. Company strategy is aligned with procurement strategy, and

procurement strategy is aligned with category strategy. (Pohl & Förstl, 2011) These alignments are harder to observe, and they are not indicated in numbers.

As mentioned in chapter 2.3, procurement performance can be measured from different angles. One approach is to divide performance into internal and external performance. Internal performance is the performance of the internal procurement functions and external performance of suppliers. Measuring saving activities in the procurement is one of the internal measures and supplier satisfaction is one of the externals. (Jääskeläinen, 2018; Pohl & Förstl, 2011) According to Weele (2018, pp. 305–306), procurement performance is divided into two elements: procurement efficiency and procurement effectiveness, which are illustrated in the Figure 6. The latter is defined as steps taken towards a target or goal which certain actions create. It refers to the relationship between the actual and planned performance of the human activity. Effectiveness relates to which extent the previously established goals have been met. Procurement efficiency means the relationship between planned and actual sacrifices made to reach certain targets. Thus, it is related to resources that are needed to accomplish previously agreed targets. Putting it to a concept, it refers to the correlation between actual and planned costs. Effectiveness is related to the targets and objectives of the procurement function. Efficiency relates to resources and their optimal use to full fill objects set to the procurement function. It focuses on human resources of procurement and different information systems.

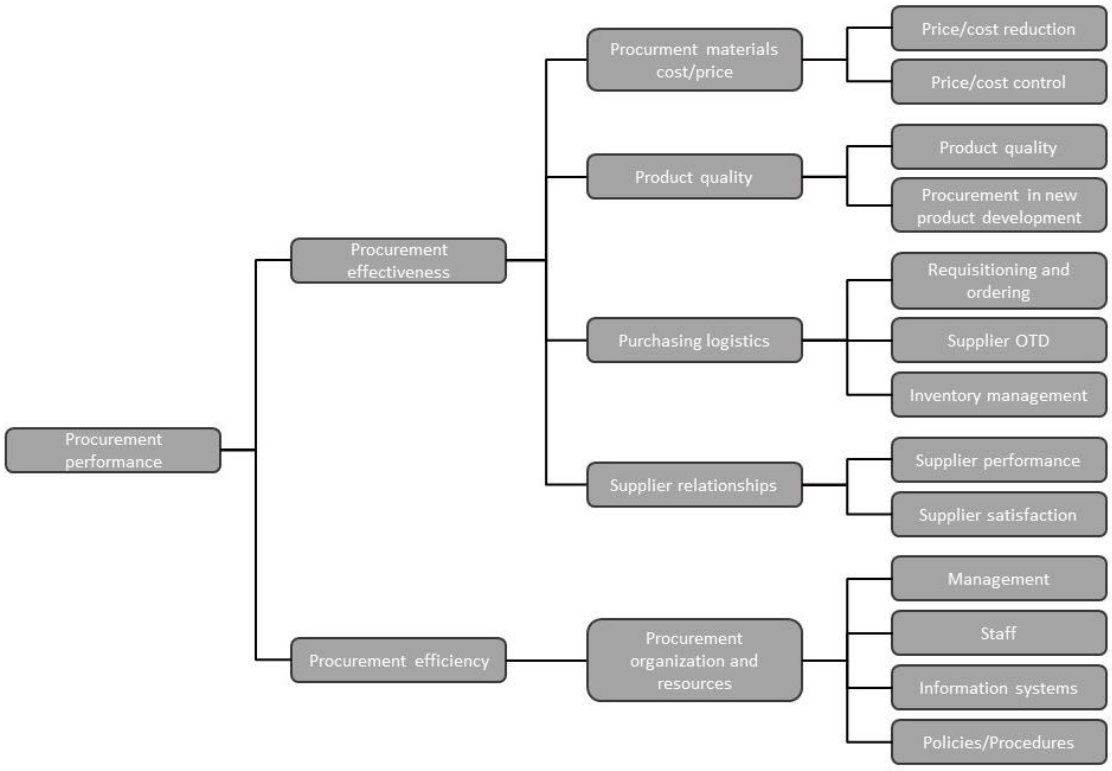

Figure 6. *Procurement performance measurement key areas* (adapted from Weele, 2018, p. 307).

Procurement effectiveness in the Figure 6, is divided into four sub-categories: procurement material cost/price, product quality, purchasing logistics, and supplier relationships. The first category, procurement material cost/price, refers to a comparison between standard and actual prices and its' objective is to monitor and control them. The price reduction focuses on the following projects and actions which intend to reduce cost on procurement. Purchasing quality relates to the quality of purchased commodities and services. The second category, product quality, is straightforward quality control on purchased products. It ensures that the delivered products are delivered and produced according to the company's specifications. The latter category illustrates procurement's involvement in product development. Purchasing logistics focuses on the efficient movement of commodities and services. The category is divided into three sub-categories: requisition and ordering, supplier OTD and inventory management. Requesting and ordering focus on controlling timely and accurate handling of purchase order requisitions Supplier relationships concentrate on improving supplier performance and monitoring suppliers. Good relationships and satisfaction between parties ensure continuity and compliance with the agreements. From a viewpoint of procurement efficiency, the single sub-category divides into organizational, informational system, and policy aspects of procurement. The objective is to develop and monitor resources that make procurement functioning, like training staff. (Weele, 2018, pp. 308–309) Jiang (2017) mentions that an

efficient supply chain is pulled by the end-user rather than the manufacturer. This reassembles the just-in-time philosophy (Leopold, 2015). Efficiency takes also into consideration the efficiency of buyers. Buyers are relying on various data sources for monitoring purchase orders and supplier performance. Monitoring supplier performance takes a lot of time and effort from the buyers. Therefore, by automating these tasks, buyers' time is saved towards other tasks. (Weele, 2018, p. 44)

Similar strategic hierarchies are present in the procurement performance measurement as in the previously mentioned purchasing strategy. Luzzini et al. (2014) in their study observed a similar hierarchy in the procurement performance measurement as illustrated in the Figure 4. The biggest differences between the hierarchies were the scope of the measurement. At the corporate level, the procurement was measured as a whole, but at the category level, interest was on product categories. This illustrates that procurement performance can be observed in different hierarchy levels as well.

One of the topics mentioned only briefly in this thesis is the measurement of quality. Currently, the case company's data cube lacks data on quality, therefore it is only mentioned briefly. Procurement quality is difficult to measure because it usually appears at the end of the purchasing process. They appear as rejected deliveries and claims towards the supplier. Weele (2018, p. 29) points out that there are usually three reasons for the poor quality: incorrect or incomplete specifications for the order, sourcing from an unsuitable supplier (incorrect sourcing decision), and the contract did not contain any penalties for poor quality. These points can be reduced by enforcing company rules for purchasing. Some ERP software can notify if, for example, a purchase order lacks certain critical criteria. The second point can be reduced by monitoring supplier capacity. This in turn needs data from suppliers' production or estimations for the capacity (Weele, 2018, pp. 29–30). Supplier capacity could be followed by showing open purchase orders and comparing total purchase orders to the previous year.

2.5 Pitfalls of procurement performance measurement

There are pitfalls regarding the performance measurement in procurement. Usually, these limitations and pitfalls undermine the whole purpose of the measurement system. Therefore, these possible pitfalls are needed to take into consideration when implementing and developing performance measurement systems.

One pitfall occurs in supplier evaluation. Purdy et al. (1994) divide supplier evaluation into direct and indirect process evaluation. Direct is the evaluation of the physical process and indirect through documentation. One of the pitfalls occurs in indirect supplier process

evaluation when the customer offers a supplier to define appropriate objectives and targets by itself, due to the supplier's better knowledge of the process. In some cases, a supplier has an incentive to put the targets too low, so that they are easy to reach. This approach undermines the management literature of objectives and goal setting. (Purdy & Safayeni, 2000) Another pitfall that occurs in indirect supplier evaluation is the reliability of information from the suppliers. This is especially true when this information is used to evaluate the supplier. The literature research has found out that when information is used for evaluation purposes, the counterpart tends to overemphasize positive information over negative ones. Information is also selected based on what looks better. (Pringle & Longenecker, 1982) The direct process evaluation relies on the assessor's ability to understand the process. Suppliers are usually well informed before an audit, and thus can prepare a snapshot of the situation. This snapshot enables to hide problematic areas. Some problems may be not visible at all in the snapshot and would require deeper investigation. (Purdy et al., 1994)

Lack of non-standard performance metrics is an obstacle in performance measurement systems. When several parties (procurement functions and suppliers) do not agree on metrics or measurement systems, the measurement becomes harder. Forslund and Jonsson (2007) found out in their studies that the lack of standardized metrics and common definitions was a general obstacle in implementing performance measurement. One area that lacks non-standard performance metrics is the OTD. In the study, lack of integration between suppliers is one element that hinders OTD development. This lack of integration emerges in the target setting for the OTD. They noticed that firms usually set the same average targets for all suppliers. This average target is not agreed upon together with all suppliers. The lack of specific targets for suppliers hinders the supplier monitoring process and might demotivate them. Thus, supplier contracts should contain specific numbers, the use of averages makes the contract work as not intended.

Procurement performance relies on the joint performance of all parties, while management is responsible towards the company shareholders (directors, upper management) and company objectives. This might create a conflict of interest between parties. They might have different goals and priorities. Forslund and Jonsson (2009) argue that if the objectives and priorities of partners are not aligned with objectives and priorities of the inter-organizational business process it is less likely that the performance measurement system is integrated between parties. These different goals and priorities are even present between business units inside the organization. Business units might have different approaches on goals settings, which creates friction on what goals to use in metrics.

A much more common obstacle of the performance measurement system is the capabilities of IT systems. Most corporate information systems are incapable of gathering non-traditional data on the supply chain (Bourne et al., 2000; Brewer & Speh, 2001). To overcome these issues corporations, need to first update their IT systems, so they can implement performance measurement systems. Most companies see their ERP system as dysfunctional and thus opt for creating meaningful performance reports with Excel. Correlation has been found between the ease of creating performance measurement reports and the frequency of these reports being done. When the report is harder to do, they are made less frequently. (Bourne et al., 2000; Lohman et al., 2004)

2.6 Need for performance measurement

The use of performance measurement varies from organization to organization. In the literature, several characteristics of performance measurement needs are defined. Bourne et al. (2000, p. 758) divide the performance measures into two main subdivisions: reviewing the strategy implementation and challenging assumptions of current strategy.

First, the need for performance measurement should come from corporate strategy. The original goal is to measure the success of the strategy implementation. Second, the feedback from the performance measurement should be used to challenge the assumptions and test the validity of current strategies. (Kaplan, 1996; Vitale & Mavrinac, 1995) Vitale and Mavrinac (1995) state in their research that the performance measurement system is an integrated set of performance indicators that glue strategic objectives into functional tasks by redirecting attention to critical outputs necessary for the strategy. Needs for the performance measurement were:

- redirect attention into what is strategically important,
- increase visibility across functions ensuring harmonized coordination and synergies,
- encourage action, and
- improve communication crucial to learning.

Challenging the assumptions of the current strategy is a crucial part of performance measurement. These two phases can interlap. Some measures can be already in place when others are under development. Additionally, a literature study by Neely et al. (2002) suggests that the performance measurement was required to:

- monitor and control,
- drive improvement,

- maximize the effectiveness of the improvement effort,
- achieve alignment with the organization strategy and targets,
- reward.

Neely (2002) states that performance measurement is for clarifying and communicating strategy. For management, the selection process of what measures are included in the performance measurement system forces management to think about what measures are necessary. This in turn has a clarifying effect. By clarifying the performance measurement system, it becomes easier to communicate the corporate strategy to different levels. It is noted that close alignment between measures and strategy can lead to more efficient and effective strategy implementation (A. Neely et al., 1994).

The corporate objective is important to the company, and it mirrors the chosen strategy and vision. Other corporate objectives might be, for example, the reduction of CO2 emissions by 2030. These corporate objectives are needed to be reflected in the lower levels of procurement. When the strategies and objectives are brought to the lower level, they direct action into the things that are necessary to the corporate's success. In the example of CO2 emission, it leads buyers and managers to consider the CO2 aspect as well. Additionally, performance measurement can derive motivation. The savings target encourages purchasing function to be sure that the commodities and services are bought on the lowest total cost of ownership (TCO) or at the best value. Thus, it works as a motivator for operational procurement. (Weele, 2018, p. 56)

There is a need for performance measurement from the perspective of purchasing and supply chain management. As mentioned in the previous chapter, there is a need to measure savings in the procurement and therefore also need to motivate employees. Several other areas have also a need for performance measurement. In the literature (Jääskeläinen, 2018; Pohl & Förstl, 2011) these needs of performance measurement are divided into two categories, internal and external measurement. Internal measurement illustrates the purchasing function's needs, such as previously mentioned savings, make-or-buy decisions, and supplier selection. In addition to these internal measurement needs, measurement is needed to monitor the performance of the procurement functions and teams. By monitoring performance, procurement efficiency can be improved. The external measurement illustrates the need for supply chain management, for instance, supplier performance evaluation. Supplier evaluation is traditionally focused on price, quality, and delivery, but now more attributes affect the supplier evaluation (reputation, capabilities, service, etc.) (Ha & Krishnan, 2008). One area of external measurement need is for increasing collaboration between the buyer-supplier relationship. Relationship

is needed to convey the common goal and vision to the parties. This is done to motivate, improve collaboration, and achieve the targets (Papakiriakopoulos & Pramatari, 2010). In the study of Jääskeläinen (2018) performance measurement was seen as a tool to make the interaction between buyer and supplier more frequent. As mentioned in the previous paragraph the performance measurement plays a critical role in aligning the corporate strategy. This is also the situation with procurement. For strategies to be aligned, a good procurement performance measurement is needed (Dumond, 1994).

2.7 Implementation of performance measurement system

Amongst the published works, there are several roadmaps and guides for implementing performance measurement systems. Parmenter et al. (2020) created a framework for implementing performance measurement systems in corporations. The framework avoids typical pitfalls in the performance measurement development process. One example of the pitfalls is that the end users are not listened to or taken into account in the implementation process. By involving the end users in the implementation process the end users are more likely to use the performance measurement system. The framework takes this into an account by introducing workshops, where users can give feedback.

Stage	Task	Project month							
		pre	1	2	3	4	5	6	post
1. Getting the organization committed to the project & training KPI team	1.1. Selling the performance measurement project to management/upper management	■	■						
	1.2 Request for the key-users to clear space from calendar for the workshops and interviews		■						
	1.3. Project lead educated themself	■	■	■		■			
2. Determine the operational critical success factors	2.1. Identify existing success factors and desired outcomes from documentation and interviews			■					
	2.2. Combine the critical success factors and perfromance measurement workshop into two-days				■				
3. Determining measures and making them operational	3.1. Refine the measures according the feedback from workshop					■			
	3.2. Hold a performance measurement gallery to drop dysfunctional and poor measures						■		
	3.3. Ask all participants to select their function performance measures from the finalized measures						■		
	3.4. Find the key result indicators (KRIs) and the performance indicators (KPIs)						■	■	
	3.5. Design the reporting framework						■	■	
	3.6. Facilitate appropriate use of the selected measures for end users							■	■
	3.7. Refine CSFs and associated measures after one year of use								■

Figure 7. *Process for performance measurement development* (adapted from Parmenter, 2020)

In the Figure 7, the implementation framework starts with the organization's commitment. The phase contains involving the upper management in the development process as well as training the staff. The next phase is about defining success factors and expected outcomes for the implementation process. Workshops are held to define the success factors and needs of functions. The last phase is about implementing and developing the performance measurement. First, based on information gathered in phase two, the performance measures are created. Then galleries are held to drop out the poor measures and the teams are asked to pick the needed performance measures. After this process, the key result indicators (KRIs) and KPIs are defined, and reporting framework is created. The final parts are about maintaining and updating the measures over time.

For maintaining and updating the performance measurement, a forum or regular meetings are required to review the developed metrics. Directors and managers who are responsible for the measurement should be present in these meetings. Meetings should focus on the measures that were developed and develop future actions for the performance measurement. In this way, the resistance from managers and directors can be reduced and the performance measurement is developed in the right direction. (Bourne et al., 2000)

Parmenter et al. (2020) argue in their study that performance measurement should be created from bottom to up. They state that the team managers and upper management need to know about the organization's critical success factors. The argument in the study is that the operational teams work with the KPIs. Supporting functions will work with the PI and RI. The different levels of performance measurement in the organization are visible in the Figure 8.

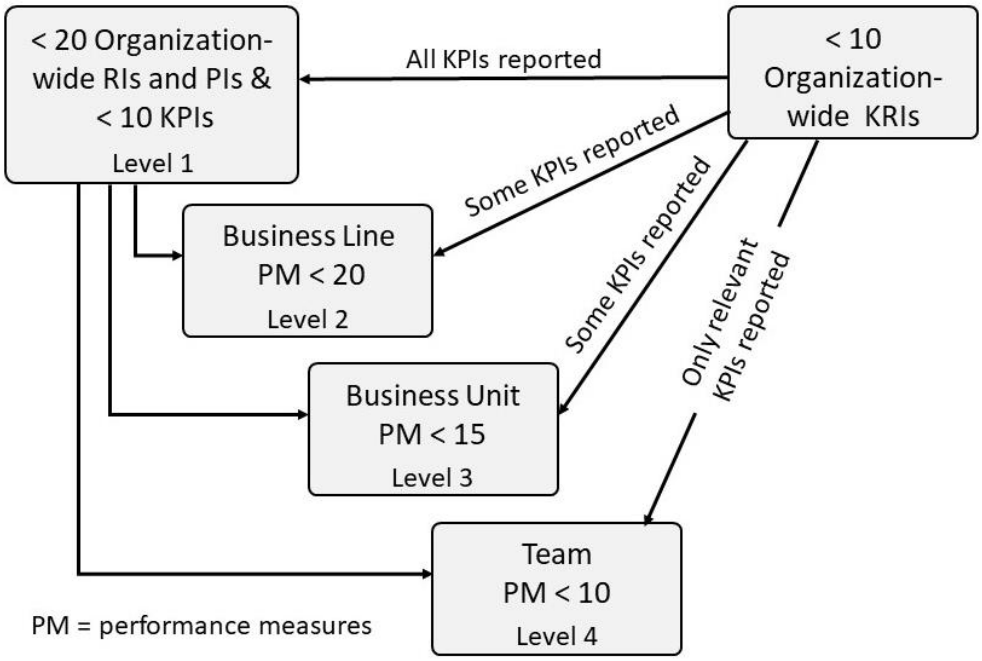

Figure 8. *Related levels of performance measurement in an organization* (adapted from Parmenter, 2020)***.***

As illustrated in the Figure 8 the business unit level performance measurement should show under 15 performance measures. Key result indicators are reported to the business unit level from the organization performance measurement level. The team performance measurement has the lowest number of measures. Only the relevant information is illustrated.

2.8 Obstacles of implementation

Bourne et al. (2000, p. 760) observed in their three case studies that it took considerable length to complete the performance measurement development. From the start of the design to the end-users actually using the system, it took 13–17 months. The design of the performance measurement system was completed usually in four months, but it took 9 to 13 months before end users actually used it regularly. Neely et. al (2002) found out in the case studies of implementing a performance measurement system that half of the cases failed to implement the performance measurement. Reasons behind the failures were categorized as fundamental strategic choices or changes in strategic focuses. The fundamental strategic choices meant that changes affect the function that is undertaking the performance measurement process. For example, restructuring of division is one. The changes in strategic focuses mean that other projects are favored over the performance measurement and therefore it is not done.

Other obstacles except for the extended development time or failure of the project were observed. Bourne's et al. (2000, pp. 760–761) found out that typically there are three obstacles in implementing performance measures:

1. Resistance to measurement during development and use,
2. IT issues during the development and
3. Lack of top management commitment.

Typical resistances are the lack of use after the implementation and general slow progress. It is hard to measure resistance because the people resisting the change tend to hide their resistance. One of the main issues is that the management might resist the implementation of the performance measures because they fear that the measures hinder their work. (Bourne et al., 2000) Thus, it is important to point out that the measures are there to help their work. Allio (2012) mentions that the management usually complains about metrics, because they oversimplify the complex process and, in that case, the upper leaders take them out of context. Other reasons behind resistance are incorrect data, high volume of metrics, or missing the intended target. Resistance may occur if the performance measurement is developed without taking the end users' opinions into mind.

As mentioned in the chapter 2.5, the second typical issues are IT issues. Corporations have different systems and legacy systems in place. This makes gathering data from different sources difficult. There are also security concerns on accessing a wide range of data, which might limit the development process. Bourne et al. (2000) mention that the corporation might be in the middle of software and database updates, which might reduce the accessibility of the data. This was the issue in the case company.

The third obstacle that commonly occurs is the lack of top management support. management might lose interest or have too much work to do to support the development process. Considering the long development process, it makes sense that the management has other things to do. The lack of support might lead to delays in the development process or the cancelation of the project. (Bourne et al., 2000)

2.9 Procurement Dashboard

One of procurement excellence's responsibilities is to develop a procurement performance measurement. Usually, these performance measurement systems are illustrated as dashboards, which contain metrics and indicators. Dashboards should have only the most important performance measures. A wide amount of data has created a situation

where a lot of metrics are implemented into the dashboards. Metrics illustrated in the dashboards can be for example realized savings, supplier performance, lead-time monitoring, and payment term monitoring. (Weele, 2018, p. 172) Visualization plays a key role in conveying the right message to the end-user. Visualizations use the human ability to visualize patterns, identify outliers, and trends. Different visualizations have been studied to (tables, bar charts, pie charts) work better in certain situations. (Abela, 2008)

2.9.1 Dashboards

At the time when a huge amount of data is available, the dashboards are getting crowded with metrics. Data is available from different ERPs across the organization, and this increases pressure to have more metrics and analysis. Although more data is gathered from various sources, a lot of companies are having a challenge with ERP's that are not integrated with the procurement information systems. This lack of data has been mentioned to be one of the issues in developing sourcing strategies. (Weele, 2018, p. 15) There can be situations where the company has no idea how much they are buying from a certain supplier or how much they buy materials. This is due because some corporations lack uniform data warehouses and systems, where all the data is combined and presented (Bourne et al., 2000). In bigger organizations, this occurs less frequently (Weele, 2018, p. 15).

A huge amount of data has led to a situation where a lot of metrics are implemented. Allio (2012) points out in their research that fewer indicators are better. Quality should be preferred upon quantity. Deciding on the correct number of measures is difficult. Stakeholders must agree about the correct measures because they steer the decision processes and the projects. The situation becomes more difficult if the company has, for example, 50 metrics already in place. It is easier to assume that the more information the better, but this leads usually to the cluttering of measurement. Considering the limited space available for measures, deciding the correct ones become important. Decision-makers need context, not more metrics to handle. If a manager or an employee cannot find a correct measurement, they come frustrated, and this may lead to neglecting the dashboard completely. (Kerzner, 2013) According to Allio (2012), dashboards should have between 10–15 indicators. Those 10–15 metrics in a dashboard should be KPIs and focused on critical success factors of the company, as mentioned in. The metrics should be simple to understand, it does not mean that they should be oversimplified. They should focus on indicating early-warning signs for the decision processes and address critical value chains. Another important factor for metrics is that they have context. Metrics should have context on them to help the leadership to see the bigger picture and

trend in the measurement. The metric should not be a single figure, for example, the number of order lines this year. This does not give any additional feedback to the end-user and left the user with more questions than answers. It lacks the trend of the order lines. It would be better to have, for example, order lines per buyer over time. This would monitor the load for the buyers, which might be useful in operational procurement. (Allio, 2012)

2.9.2 Visualization

Visualization is a key aspect of good performance measurement. It is considered in the design and implementation phases of the performance measurement. It is directly visible to the end-user and can be a powerful tool for how the performance measurement is designed, implemented, and used (Jääskeläinen & Roitto, 2016). With visualization human ability to visualize patterns, see trends, and identify outliers is used to help understand data better and faster. Good visualization can replace cognitive calculations with a simple interface and improve understandability, memory, and decision making. This way data is made more accessible to a more diverse audience, and it can engage exploration and analysis of data. (Heer et al., 2010) Highly nonhomogeneous and noisy data can be coped with visual data exploration easily (Keim, 2002). With good visualization, data exploration is faster compared. The challenge is to find the right visualization for the right data.

Visualizations can be categorized in several different ways. Lengler and Eppler (2007) categorize visualization into several categories, the most important one for the thesis being data visualization and information visualization. The data visualization contains the familiar quantitative formats: bar charts, pie charts, line graphs, etc. They give the user a bigger picture of the data and answer the questions “how much”, “how”, and “why”. Information visualization is used to amplify cognition. Visualizations in use are, for example, tree maps and semantic networks. The end-user can change the visualization. Keim (2002) divides visualization into dynamic and interactive categories. Dynamic visualization is updated automatically without the user’s input. The interactive is updated by the user. Another approach is to divide visualization into functional and aesthetic information (C. Chen, 2010). Functional information visualization intends to present the data as accurately and truthfully as possible. It has little focus on aesthetics. Aesthetic information visualization on the other hand focuses on aesthetics with the cost of functionality. (Lau & vande Moere, 2007) The aim is to present visually pleasing charts and graphs. In the literature, there is a debate about which methods to use. Some authors propose to

use both functional and aesthetic methods together (C. Chen, 2010; Lau & vande Moere, 2007).

Graphs and charts can be divided into several categories. For the graphs and charts, Heer et. al. (2010) divides them into several categories:

- time-series data:
    - stacked graphs,
    - horizon graphs,
    - line charts,
- maps:
    - graduated symbol maps,
- networks:
    - matrix views.

Stacked graphs, small multiples, and horizon graphs are categorized under time-series data. In the time-series data, the values are changing over time. These types of graphs are the most commonly used (Heer et al., 2010). Stacked graphs enable the user to visualize which entities the total value consists of overtime. Stacked values are not suitable for negative values and are only useful for values that can be summed up. Column charts are suitable for presenting a few items. Horizon graphs increase the data density in a graph while preserving the resolution. Horizon graph has a learning curve to read it, but it is more effective on small charts compared to standard plots. (Heer et al., 2010) Graduated symbol maps are visualization where values are visualized over a map. For example, in procurement purchase orders by countries can be illustrated on a map. Matrix view is a good way of presenting complex networks and comparing them. The matrix enables users to explore network connections by drilling down on values, grouping, and reordering of values. (Heer et al., 2010)

Abela (2008) approaches chart categorization with four main categories based on end-use: comparison, distribution, composition, and relationship. This is illustrated in the Figure 9. The diagram is made to help choose the right chart for the right situation. There are similarities to Heer et. al. (2010), where column charts and line charts can be found categorized under the time-series data. Matrix/table is also situated in Abela's (2008) diagram to the comparison category with the ability to compare many items.

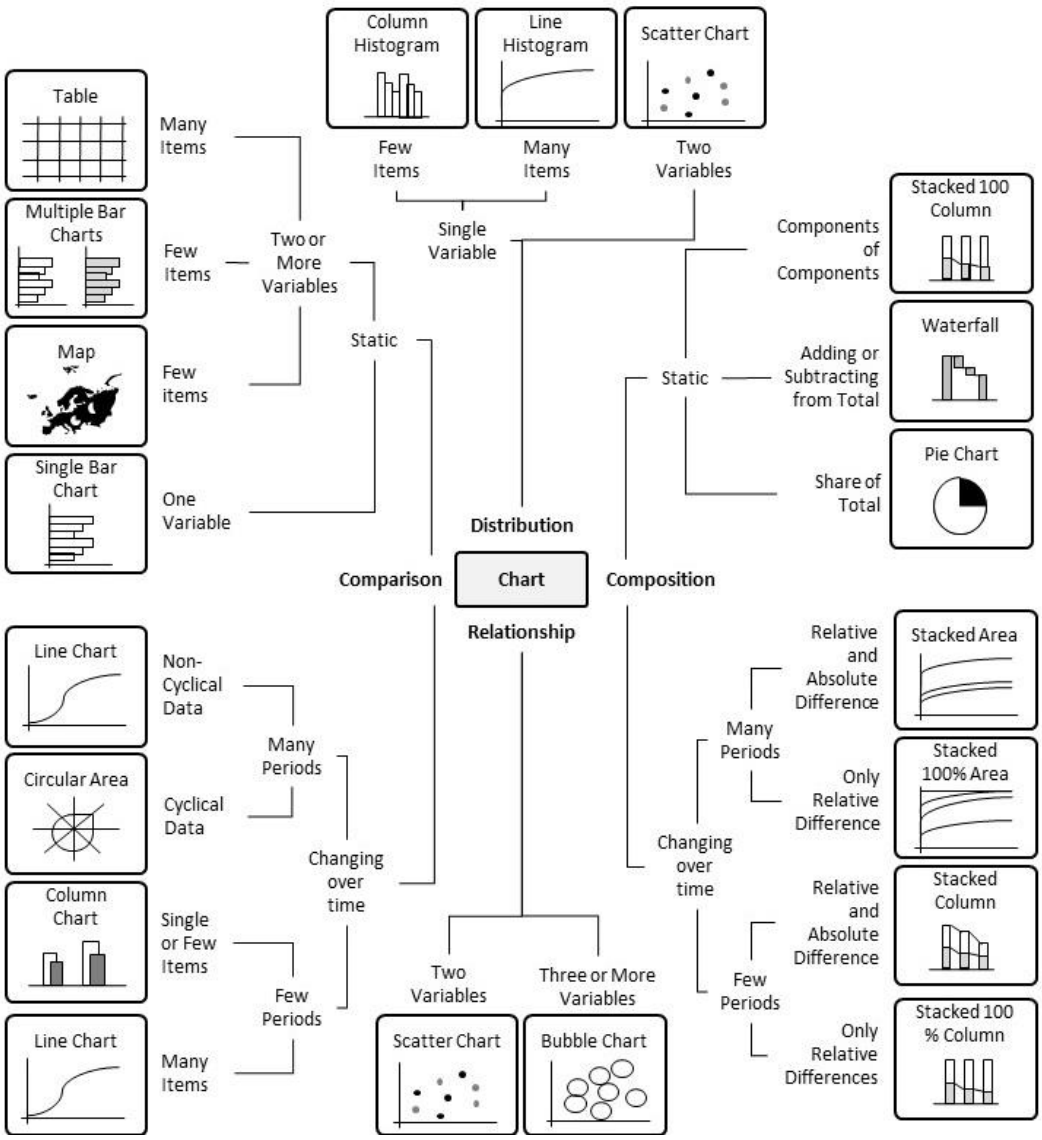

Figure 9. *Chart categorization* (adapted from Abela, 2008).

As visualized in the Figure 9, the chart can be divided into four main categories, which each have additional sub-categories. Composition gives the user an easy look at total numbers with static or continuous time. Horizontal bars charts are more effective in showing the sub-totals of which the total value consists (Järvenpää & Dickson, 1988). If there are many periods, lines are used because trends are easier to recognize. Column charts that have columns close together are also suitable for visualizing trends. (Shah et al., 1999) In comparison, if there are only a few variables, horizontal columns are used. This is because people tend to overestimate vertical bars compared to horizontal bars (Järvenpää & Dickson, 1988). In the Figure 9 relationships between variables are illustrated with scatter charts. Lewandowsky et al. (1989) found out in their study that illustrating the variables in scatter charts with different colors, shapes, or letters, increases understandability. Therefore, it is useful for showing numerous variables.

The literature has mixed results on the use of colors in the charts. Based on literature research by Abela (2008) colors can be helpful for the less advanced audience because they can direct attention towards problematic areas. For a more advanced audience, colors can be distracting because they can distract attention. Colors should be used to highlight certain areas, but not too much.

3. RESEARCH METHODOLOGY

In the chapter, the research methodology of the thesis is presented. After the methodology, the research strategy and timeline are stated. Next, the literature background is connected with the empirical studies. Finally, the used data gathering methods are introduced.

3.1 Research methods

The time horizon for the thesis is a *cross-section*. This is due to the limited time and resources available. In cross-sectional research, some phenomenon is researched at a specific time and are thus time constrained. (Lewis et al., 2019, p. 212) *The case study* is used as a research strategy for the thesis. In the case study, research can focus in-depth on the case and retain a holistic and real-life viewpoint (Yin, 2018). In this thesis, the case company and its procurement are used as a case. To better understand the case fully, data is gathered from qualitative and quantitative sources. Thus, the methodological choice for the thesis is the *mixed method*. In the thesis double-phased sequential research design is used, where both qualitative and quantitative data is gathered but the quantitative data is used as a support for qualitative data. Qualitative data is gathered from interviews, workshops, and the case company's internal materials. The company's internal material is used for defining company strategies and objectives. Quantitative data is gathered from the case company's data cube.

The research philosophy affects the assumptions and beliefs of the research. The assumptions and beliefs shape the understanding of the knowledge. With this in mind, the *pragmatism* philosophy was chosen. Kelemen et al. (2008) propose that in pragmatism the concepts are only relevant if they are relevant to some action. Multiple types of methods are possible and preferred in pragmatism. Although multiple methods are possible the methods that are credible, reliable, and relevant are used for the research (Kelemen & Rumens, 2008). In pragmatism, no single method can give a complete picture of the situation. In the pragmatical approach, research starts with a problem (Lewis et al., 2019, pp. 150–151). In the thesis, it is the objective of the thesis to create harmonized and strategically unified dashboards for the procurement. A practical solution is needed to solve the company's problem. This solution could also support future practices.

Induction was chosen as an approach for theory development. In the inductive approach, theory follows data. In induction, data is gathered on a specific theme. This data is used

to create a new or modify the existing theory. The new theory is tested with additional data. The fundamental idea behind induction is to first get a basic idea of the topic, analyze the gathered data, and make a theory from it. (Lewis et al., 2019, pp. 153–155) In the thesis' the Parmenter's (2020) framework for *implementing performance measurement system* and Neely's et al. (2000) framework for *performance measurement system design process* are modified and used. Additionally, Abela's (2008) *visualization framework* is used to determine the type of the measures. The data to test the frameworks is gathered from the workshops, interviews, personal meetings, internal documents.

All the interviews in this thesis are conducted as *semi-structured interviews*. In a semi-structured interview, the interview starts with a pre-determined theme and with some key questions to guide the interview. The idea is to gather information around the theme and analyze the common topics from the interviews. When the semi-structured interview is used with an inductive approach, the interview themes will naturally evolve around the emerging topics from the analysis of the interview data. The selected research methodologies for this thesis are illustrated in the Figure 10. (Lewis et al., 2019, p. 438)

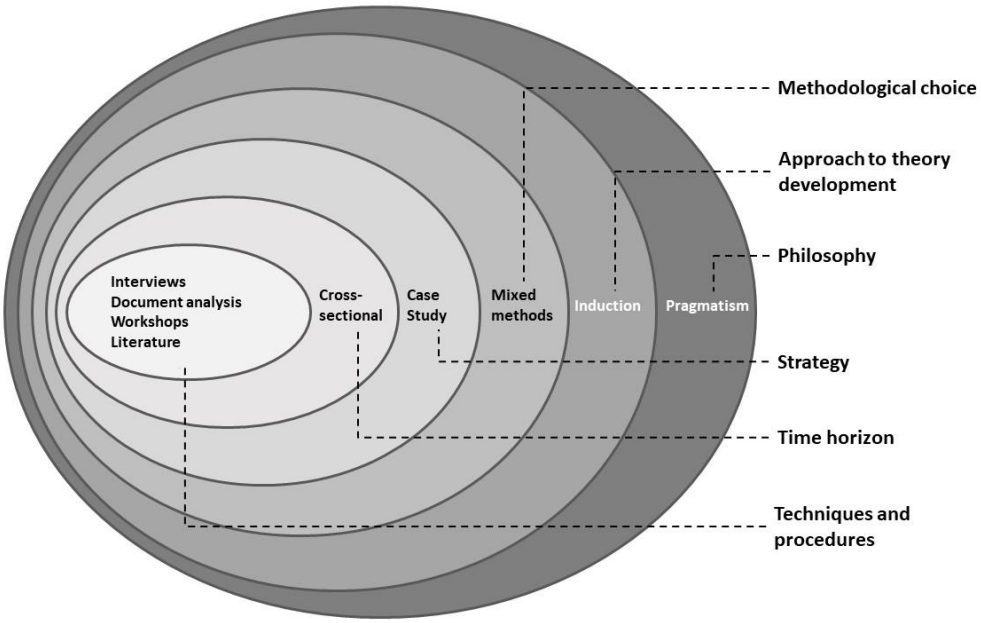

Figure 10. *Research methodologies of the thesis* (adapted from Lewis, Saunders and Thornhill, 2019, p. 130).

The research onion created by Lewis et al. (2019, p. 130) is used to represent the chosen methodologies. In the center of the Figure 10, the data collection and analyzing methods are illustrated. On the right side, the different methodological and research concerns are illustrated.

3.2 Research process and timeline

The research process starts with the literature research, about procurement and performance measurement around procurement. This is done to gain knowledge about the topics at hand. The background of the case company is observed and studied to form an understanding of the practices in place.

The first part of the research is to know the current situation of the performance measurement in the case company. This seeks to partially answer the first research sub-question (SQ1) about the current situation. The focus will be on understanding the current practice and extend of procurement performance measurement.

The second part of the research is about defining the need for performance measurement of procurement stakeholders. It will answer the second sub-question (SQ2) about the needs of performance measurement. The semi-structured interviews are at the center of answering the question. Neely's (2000) framework for performance measurement design is used as a frame in the interview. The questions are focused on the current situation, defining the need for performance measurement, identifying needed performance measures, and future development needs.

The third stage of the research is to build and define the individual performance measures. The created performance dashboards are based on the information gathered from the literature and interviews. The dashboards are developed further based on the feedback from the workshops. Frameworks are used to define the ideal measure in the dashboards. For example, the strategic aspects of the measures are assessed to ensure that targets and activities of operational procurement are aligned with the strategic procurement objectives. Findings from creating the measures will answer the third sub-question (SQ3). The development will continue after the workshops, but it is out of the scope of the thesis. In this stage,

These three stages help to answer the main research question of this thesis (Q1): *"How to develop a procurement measurement system in a company operating in a manufacturing project business?".* Research and the implementation process will be the key reference source for answering the research question.

The timeline for the thesis is illustrated in the Figure 11. The timeline starts with planning research and a review of the literature. These steps are crucial for the success of the thesis. The literature review gives information on the topics from which it is possible to form the interviews.

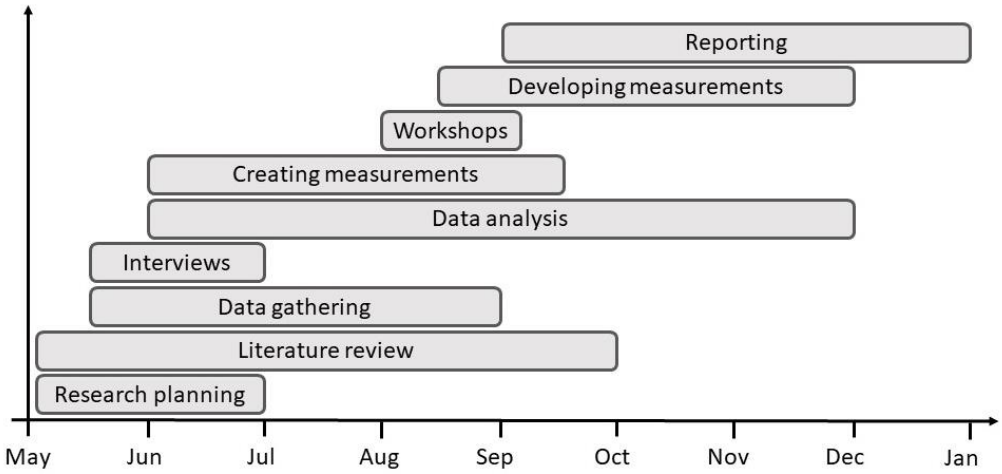

Figure 11. *Timeline for the thesis.*

The timeline for the thesis is affected by the summer holiday season in the case company. The holiday season created an urgency to finish interviews before July. The reason was to have sufficient time for data analysis and to create measures before the holiday season. Therefore, most of the interviews are done before the start of July, which is illustrated in the Figure 11. After the holiday season, the workshops were established to gather development aspects and to involve key users in the process. After the workshops, all the necessary data is gathered, and the remaining time is devoted to the writing of the thesis.

3.3 Connecting literature background to the empirical study

In the empirical part Neely's (2000) requirements for the performance measurement system design were introduced. These requirements are used when creating interview questions, which are illustrated in the Table 3. The most common obstacles in the implementation process and pitfalls of performance measurement are also considered when creating interview questions about the risks of the implementation process.

For the performance measurement implementation process, modified Parmenter's (2020) development process is used. The same development process is used for all three dashboards (business unit, category management, team management). As mentioned in the 4.4 chapter, the refining process of the Figure 7 is removed, selling of the KPI project to the owners is removed because the owners were already committed to the project. Task 2.2 from Figure 7 is removed and replaced by the interviews, and finally, the maintenance section is removed due to the limited scope of the thesis. In addition to Parmenter's (2020) implementation process, Neely's (2000) performance measurement

system design process is used. The framework is used in the interview stage when refining the questions. In addition to interviews, the process is used in defining the measures.

In the metrics design process, the visualization framework by Abela (2008) is used to determine the right visualization for different situations. For determining the right number of metrics in the dashboards, Parmenter's (2020) related levels of performance measurement in the Figure 8 are used in addition to Allio's (2012) guidelines for dashboard design, scope, use, and impact.

3.4 Data gathering

Data were gathered through semi-structured interviews, workshops, private meetings, from the case company's data cube, and the case company's internal material. The *qualitative primary data* was gathered through interviews to get an understanding of the current situation, preferred performance measurement, and future development needs. Primary data was gathered also from the company's internal material. For example, strategies and objectives of the procurement were gathered from internal sources. *Secondary quantitative data* is gathered from the case company data cube to create the performance measurement dashboards. As illustrated in figure 6, the data cube gathers information from multiple data warehouses, ERPs, and other data sources. This data was then modified, and features are added to it to get it more suitable for the business line in hand. Modifications are possible to a certain extent and thus some features are not able to be implemented. Secondary data was also gathered from the legacy performance measurement system. Information was gathered outside of the workshops and interviews through private meetings. Private meetings were placed when unexpected development needs rose, or clarifications were needed to the performance measurement. Private meetings were held through Teams or at the office. Notes were written about the meetings.

3.4.1 Semi-structured interview

The interview round was done to *gather qualitative primary data* for implementing the performance measurement dashboards. The interview was structured with questions about the current situation, possible development paths, preferred metrics, measurable things, and possible risks in the development process. The structure follows the framework illustrated in Neely's (2000) research. *Semi-structured* interview approach was chosen because of the high answering rate and relaxed environment, which was hoped to increase objectiveness and discussion. Almost all requested interviewees were able to

attend the interview. Thus, the scope of the interview was as intended. All interviews were recorded, making it possible to return to listen to the interviews again. Permissions were asked beforehand for the recording. The interviewees and information about the sessions are visible in the Table 2.

Table 2. *The interviewees of the first interview round.*

Interviewee	Title	Date (dd.mm.yyyy)	Duration (h:mm)
P01	Vice-president, Operational Excellence	1.7.2021	0:46
P02	Director, Supply Chain	18.6.2021	0:43
P03	Director, Supply Chain & QHSE	2.7.2021	0:48
P04	Director, Procurement	30.6.2021	0:55
P05	Director, China Supply Chain	2.7.2021	0:36
P06	Category manager	2.7.2021	0:55
P07	Category manager	22.6.2021	0:48
P08	Manager, Global sourcing	29.6.2021	0:59
P09	Manager, Production Procurement & Supply Chain	30.6.2021	0:59
P10	Purchasing Manager	1.7.2021	0:32
P11	Trade coordinator	28.6.2021	0:47
P12	Trade coordinator	29.6.2021	0:38

Due to the COVID-19 pandemic, all the interviews were done with Microsoft Teams. The software also allowed fluent screen sharing, which was useful when presenting current procurement performance measurement and the interview questions. The interview questions were planned beforehand, ensuring that all critical topics were covered. The interviewees got the questions before the interview to save time. The questions are illustrated in the Table 3. Requests were also made to the interviewees to deliver the current performance measurement before the interview. The intention was to save time around the topic of the current situation. The time saving was urgent because interviews had to be done before the holiday season. The participants had already a heavy workload because of the upcoming holidays.

Table 3. *Interview questions.*

Number	Question
Q1	For what purpose do you use the procurement metrics of your business unit? To which stakeholders do you supply/present the metrics? What kind of actions do the measurement results typically lead to?
Q2	What is the current situation regarding the procurement metrics in your business unit?
Q3	How do organization's and business line's strategies and must-wins appear in your business unit's current procurement metrics?
Q4	How the objectives of your/your team activities are visible in current metrics? How would you develop them?
Q5	What are the challenges associated with your business unit's current procurement measurement? Which meters don't work? Why?
Q6	How would you develop existing procurement metrics in your business unit? On what purpose would you like to use the measurement for?
Q7	Which current procurement metrics are useless? For what metrics there are no use?
Q8	For which different units can there be dashboards in your business unit?
Q9	What do you consider to be the most important measurable things in your business unit?
Q10	Which metrics do you consider important in your business unit? What metrics should be included in dashboards? Why?
Q11	Which metrics are good?
Q12	What kind of thing should the procurement measurement dashboards form as a whole? What kind of connections should the metrics have?
Q13	What kind of manual reports could be useful to automate?
Q14	What do you think are the biggest risks in developing business unit measurement and dashboards?

The interview questions follow *the performance measurement design principles* illustrated in Neely's (2000) research. The first intention of the interview is to find out what measures are required and what is the current situation. What information do managers need to manage the business? The second intention is to determine the most important measures. The third intention is to find the purpose for the measures. The fourth intention is to find out additional metrics. The first two questions, in the Table 3, are intended to

find the answers for the current situation in the case company. The first question focuses on the purpose of the measures. Q3 and Q4 are trying to figure out how the organization's strategies and objectives are visible currently and how they could be developed. Q5, Q6, and Q7 are about the challenges in the current performance measurement system and how they should be developed. This is in line with the Neely's *modified performance measurement system design principles* that take into account the strategies and policies of the organization. Q8 asks the interviewees' opinion on which entities should be measured in the interviewees' business unit. Questions 9 and 10 intents to find what the most important measures are. Design and connections of the measures are assessed in the Q11 and Q12. The Q12 asks about the bigger picture about the performance measurement and what connections there should be between metrics. The final question, (Q14) intends to find possible risks and pitfalls in the development process. One of them is the maintenance of the performance measurement mentioned in Neely's research.

3.4.2 Workshops

After the interview round and the implementation phase of the draft dashboards, two workshop meetings were held. In the first workshop meeting, results from the interviews were shown and drafts of the dashboards were presented. The idea of the first workshop meeting was to gather further development ideas and needs for the dashboards. The second workshop round was about finalizing the logic behind the metrics and fine-tuning the performance measurement. Workshop meetings were held with three groups of key-user representatives from different business units. Groups were divided into business unit directors, category managers, procurement team managers, and buyers. This approach ensures that the workshops don't grow too large, and everybody has time to present opinions. Every business unit under the groups was included in the workshops so that business units could make suggestions about the performance measurement. Due to many participants being in different locations and countries, the workshops were held online. Notes were gathered from the workshops to document the proposed development ideas. Participants of the business unit performance measurement workshops are presented in the Table 4.

Table 4. *Business unit performance measurement workshops.*

Workshop 1 12.8.2021	**Title**	**Workshop 2 26.8.2021**	**Title**
B1P01	Director, Supply Chain	B2P01	Director, Supply Chain
B1P02	Director, Supply Chain & QHSE	B2P02	Director, Procurement
B1P03	Director, Procurement	B2P03	Director, China Supply Chain
B1P04	Director, Delivery Operations	B2P04	Senior Manager
B1P05	Director, China Supply Chain	B2P05	Purchasing manager
B1P06	Purchasing manager	B2P06	Director, Global Sourcing & Category Management
B1P07	Director, Global Sourcing & Category Management		

Business unit performance measurement workshops focused on illustrating the draft dashboard. The proposed measures were based on corporate strategy, interview results, and individual needs of the business units. In the workshops, these measures were iterated. Between and after the workshops, the dashboard data was validated and the problems were assessed.

Category management workshops were similar to business unit workshops. The difference was that the workshops included mainly category managers and the workshops focused on performance measurement of category management. All category managers were present in at least one workshop. The participants for the category management workshops are illustrated in Table 5.

Table 5. *Category management performance measurement workshops.*

Workshop 1 13.8.2021	**Title**	**Workshop 2 26.8.2021**	**Title**
C1P01	Category manager	C2P01	Category manager
C1P02	Category manager	C2P02	Category manager
C1P03	Category manager	C2P03	Category manager
C1P04	Director, Global Sourcing & Category Management	C2P04	Manager, Global sourcing
		C1P05	Category manager
		C2P06	Director, Global Sourcing & Category Management

Team management workshops were held for more participants because key users from all business units were invited. The workshop focused on team management aspects of performance measurement and how to align the targets and objectives of operational procurement with strategic procurement. The participants of the workshops are illustrated in Table 6.

Table 6. *Team management performance measurement workshops.*

Workshop 1 16.8.2021	**Title**	**Workshop 2 26.8.2021**	**Title**
T1P01	Manager, Production Procurement & Supply Chain	T2P01	Manager, Production Procurement & Supply Chain
T1P02	Purchasing Manager	T2P02	Purchasing Manager
T1P03	Procurement manager	T2P03	Procurement manager
T1P04	Development manager	T2P04	Development manager
T1P05	Senior manager	T2P05	Senior manager
T1P06	Manager, Mill sites procurement	T2P06	Manager, Mill sites procurement
T1P07	Senior manager	T2P07	Senior manager
T1P08	Sourcing manager foundry	T2P08	Sourcing manager foundry
T1P09	Purchasing manager	T2P09	Purchasing manager
T1P10	Director, Global Sourcing & Category Management	T2P10	Manager, purchasing
		T2P11	Purchasing manager
		T2P12	Director, Global Sourcing & Category Management

Before every workshop, the participants got the link to the updated draft dashboards. Therefore, participants had time to analyze the draft dashboard and do data validation. Feedback from the dashboards was also received through private messaging and meetings, which were noted.

4. RESULTS AND ANALYSIS

Results from the interviews, workshops, materials provided, and individual meetings between the end-users are analyzed. An analysis is done to form an understanding of the research questions and objectives. Results of the development process are illustrated and the decisions behind them are explained.

4.1 Current state of performance measurement

To answer the first sub-question SQ1 an analysis of the case company's current procurement performance measurement was done. The case company has legacy performance measurement software in use, which gathers information from ERPs. The end-user has been able to create their own dashboards and filter the data as preferred. Although filtering has been a useful feature, it has created problems where different performance measurements are not comparable with each other. Even though the underlying data is the same for every user, the possibility to create and filter individual dashboards has made it difficult to cross-compare functions. One could use different assumptions when filtering data and end up with different values. There was also a possibility to include only those values that make the function look better. Another downside for the possibility to filter the data was that there was no central place to find the specific dashboards for business units, because of the assumption that everyone filters the data by themselves. The legacy software had also limitations regards of the filtering. If there was a need to filter other entities or areas besides the current one, the previous filters had to be deleted. This increases manual work and the possibility of errors.

The case company was in the middle of data warehouse development and new ERP system implementation. The case company has numerous legacy ERP systems in place from mergers and acquisitions over the years. This created some issues in the development process when the old legacy ERPs were used, and they did not communicate with the data cube. The issues led to reduced accuracy in some regional areas or excluding some regions totally. India was a region that had to be excluded due to a lack of matching ERP systems. The data cube was in the development phase at the time and did not have complete coverage of the procurement metrics. However, the impact of this was mitigated by inputting manual data from users. This was done as a temporary solution until the data warehouse was complete enough.

To tackle this issue the case company has created a cloud-based data cube, which gathers data from current and legacy ERPs as well from other sources. This data was then presented in the dashboards. This is illustrated in the Figure 12 below. By using the data cube as a source for the dashboards, ensured that the underlying data was the same for all dashboards. Thus, different business units can be compared side by side and the logic behind the calculations is similar and clear to all users.

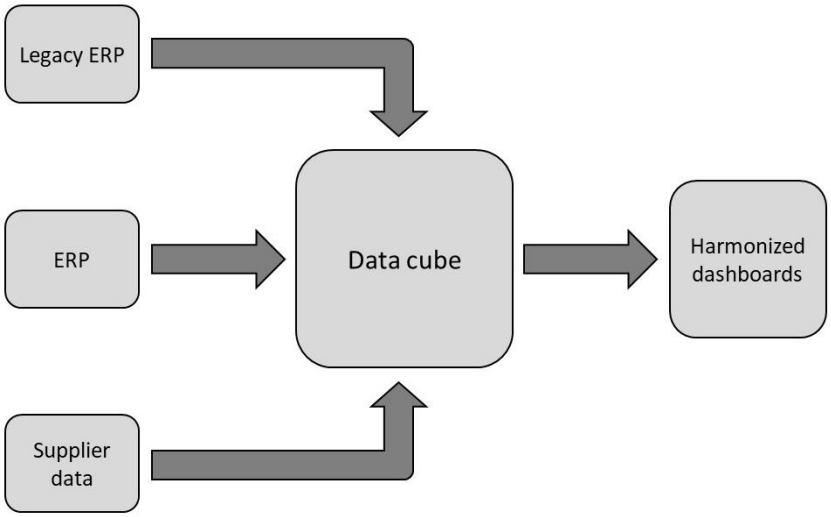

Figure 12. *The use of a data cube to harmonize dashboards.*

Lack of coverage of the legacy performance measurement system had led to a situation where end-users have had to calculate the absent metrics manually. Manual calculations have been time-consuming, especially if the raw data needs a lot of filtering and the data is gathered from several sources. Errors are prone to happen which leads to misleading results. Manual calculating favors some measures over others and might end up hurting strategically important measures that are not used in daily work. The data cube illustrated in the Figure 12 can collect data from a wider area across the organization and therefore achieve wider coverage. According to the developer team manager, this was one of the key drivers to the creation of the data warehouse.

When asked about the current situation of the performance measurement in the interviews, all but two interviewees pointed out that the current state of the performance measurement is in a good state in the case company. Several interviewees answered the Q2 in Table 3, that the current performance measurement has a wide scope, and they can find all the necessary metrics that they need. Even though the feedback was mostly positive for the second question, there were situations where the legacy software falls short. Those were especially in the business unit -level, where a lot of manual work for creating the metrics is needed. The Director of the supply chain (P02 from Table 2) mentioned that he is spending over two hours summarizing data to create performance

measurement for functions under his business unit. The second area where the legacy system falls short was when measuring CCC. When measuring CCC it would be good to see the forecast for the becoming months. The legacy software does not include forecasts with the open purchase orders. This was pointed out by the procurement director (P04, Table 2):

"When taking CCC numbers from the legacy system, the forecast is almost nonexistent. It is not good. We have the open purchase orders option available. The legacy system does not cope well when trying to forecast CCC because it tries to find receipt dates for open purchase orders. Now when looking at the CCC we are looking at the rearview mirrors".

The issue pointed out by the director illustrates the need for the forecast in the measures. It also illustrates the limitations of legacy software. Availability to forecast the CCC is needed for the director to see where the business function is going with the measure. The forecast would work by identifying purchase orders that are not received. Using the open purchase orders, the supplier can be identified by the country and therefore identified as CCC or not. With the forecast, the user knows before the purchase orders are received where the function is heading on CCC values. Therefore, for example, the director can use the information to do decisions or convey the message to the upper management.

Although several interviewees mentioned that the coverage of the legacy performance measurement was good enough, the one issue that arose from the interviews was the manual work required to build the measures. The director of the supply chain (P02 in Table 2) stated the following for the interview question two (Q2 in Table 3):

"We have good coverage. I would say we are measuring pretty much what we need to measure. And we can develop more if we want. The main issue is how much manual work do we need to add to get those KPIs".

The director pointed out the balance of the manual work for creating the measures. Not every process can be automated, or the value-added was not worth the time. In the case company, the different ERPs that did not communicate together increased the manual work required for building measures. The increased manual work was also noted by the vice-president (P01, Table 2) in the interviews. There was a situation where the director had to gather data from several sources to get the needed measures. The data had to be filtered and united so that the data is coherent with each other. As imagined, this takes time. The director of the supply chain (P02 Table 2) spent several hours creating the different measures at the start of each month. One aspect that increased the manual

work was the limitations in the legacy ERP system. On an operative procurement level, manual work was required to see the performance of the teams. The legacy system had filters for different organization levels, but lacked some at the team level, especially with newly established teams. Thus, a user had to manually select the team members and connect them with the right values.

ERP difficulties were not the only ones that increased manual work. The purchasing manager (P10, Table 2) mentioned in the interview that the current OTD measurement had difficulties. When the order was signed to the system the times were not inputted right. This reduced the accuracy of the OTD and increased the manual work required to get accurate results. In the implementation process of the performance measurement, the balance between manual work and automation had to be considered numerous times.

4.2 Need for performance measurement

The importance of performance measurement is visible from the interview results. One key aspect of performance measurement is the need for monitoring the functions and reporting the measures to the stakeholders. These needs were investigated in the first question of the interview in Table 3. The first question of the interview has a mentioning about to whom the interviewee is reporting or presenting the performance measurement. The supply chain director (P02, Table 2) mentioned the following:

"We have the top KPI for the location unit, followed by supply chain function which is under my responsibility. In the supply chain, we have 5 different functions. Each one of those has its own scorecards. Then it goes down to all employees. We have this alignment or redline from the business line from top to bottom. So, we are working with the same priorities".

This illustrates the complexity and need for performance measurement. Especially directors have several stakeholders to whom they are reporting. They usually also have responsibility for a certain area and need to monitor their performance. The director of the supply chain (P02, Table 2) has responsibility for the supply chain function and needs to monitor its performance and take corrective actions if the results are not going towards an agreed strategy. The performance measurement is also needed for aligning the corporate strategies with the strategic- and operational procurement. As the director mentioned that the case company has this alignment going from the top of the business line to the bottom. Performance measurement is needed for conveying the corporate strategy all the way to the operational level. The conveying of the message also goes the other

way around. The results of functions are transferred to the corporate level with the performance measurement. In this case, the supply chain director reports the measures to the location unit.

In order to be sure that the right message is conveyed to the different levels of the organization, measures need the be re-evaluated from time to time. The KPIs are re-evaluated in certain intervals to fit the corporate strategy and objectives. In the case company, the interval is one year, as the director of supply chain (P02, Table 2) mentions:

"At the end of each year we define the next year KPIs. Some change from year to year but some main targets always appear on the scorecards, like savings and COGS reduction".

Every year the KPIs are re-evaluated to fit the current long- and short-term strategy. The needed performance measures vary throughout the years, but the main KPIs appear over the years. Based on the interview results these main KPIs are savings, cost of goods sold (COGS), CCC, and OTD.

The need for the performance measurement rises also from encouraging employees towards a certain action. The vice-president of operational excellence (P01 in Table 2) mentioned the savings as a motivator for the procurement:

"The main reason why we measure the savings is that it works as a motivator for the procurement".

Savings works as a motivator for the operational procurement similarly as salesman's motivation is the provision. As mentioned in the theoretical background, with the measuring of savings the procurement function can be sure that the commodities and services are at the lowest TCOs or the best value. Not to forget, that the important factor for encouraging savings is also that with every euro saved the same euro shows instantly in the profits. Therefore, the company has an incentive to pursue savings to gain more profit for the shareholders. Increasing savings is more effective than making a new sale of one euro because direct and indirect costs are reduced from it.

4.3 Supporting strategic and operational procurement

When asking about which different units would need dashboards (Q8, Table 3), and what kind of connections between dashboards should there be (Q12, Table 3) the answers had a similar pattern regardless of the interviewee. Answers are illustrated in the Table 7 below.

Table 7. *Interview results on performance measurement supporting procurement*

Interview results on questions 8 & 12
Data has to be harmonized, good quality, and open regardless of what organization-level end users are
Data and performance measurement has to be easy to understand and use
Strategic procurement should have separate dashboards but be connected to operative procurement dashboards.

The interview results pointed out the need for harmonization, quality, and visibility across the procurement function. The importance of visibility was pointed out by the director of supply chain & QHSE (P03, Table 2):

"In daily work, for a buyer, there is no direct need to see must-win metrics constantly. But there has to be a possibility to that person to see those metrics so that the person can connect the operative work to it and how it affects the must-win metrics".

This is consistent with the previous chapter's vice-president's comment about the savings impact on buyers' motivation. It was stressed out by the interviewees who worked in operative procurement that it is necessary to see the results of their own work and to see how it affects strategic level metrics. As mentioned in the "Current situation" -chapter, current legacy systems lacked indicators on the operative procurement level.

As illustrated in the Table 7, there is a need for the performance measurement to be easy to use and to understand. The logic behind the metrics has to be clear to the end-user. If this is not met, the data could be interpreted incorrectly. This was pointed out in an interview by the vice-president (P01 in Table 2) who mentions the following:

"It is important to me to have contexts in the metrics. The measurement has to be easy to understand in context because they are presented in a short time window."

Some end-users were not familiar with the introduced new logic behind the performance measurement. This created confusion and misunderstanding. In the workshops, these logics were clarified, and the explanatory sheet was requested where all the metrics and their logic behind them were explained. This explanatory sheet is illustrated in the appendix. To reduce the confusion, labeling of must-wins metrics and operative metrics was also proposed. It was seen that with this approach it would be easy to identify the metrics from one another. In further discussions this idea was scrapped because of the maintenance it would require.

One of the most confusing aspects was the change to the calculation logic of used money in the procurement. The users were confused with the new logic, and it illustrated the

need for an easy-to-understand performance measurement system. The new logic was introduced in the data cube implementation process that the case company was working on. The new logic made available three different options of measuring used money compared to the previous one. By default, the legacy system is calculated based on received purchase orders and called it as spend. This created confusion because names and the logic were different in the new system compared to the legacy system. The three new options were:

- spend, based on received invoices,
- purchase order amount based on the order date, and
- received a purchase order, based on the receipt date of the purchase order.

Realization of the different calculating logic behind used money was an important result of the workshops. Based on which used money was calculated (spend, purchase order, receipt order) it had a significant effect on results. As mentioned, the spend option allocates the used money based on received invoices. In this case, it only indicates the money that has been really spend. The problem with the spend is that it lacks forecasting. If the used money is calculated with purchase order based on the order date, it has forecasting availabilities because it includes the open and received purchase orders. The problem is that those open purchase orders can include some orders that are never realized or in other words are canceled. Thus, the purchase order shows the ordered amount but does not take into consideration is it canceled or not. Thus, relying only on purchase orders based on the order date can give too opportunistic values. The final option is the received purchase order. It calculates the purchase orders that are received and have received dates marked in the system. Thus, it does not have the open purchase orders included and it shows almost the same as the spend does. One option that the legacy software had was to include open purchase orders with the received purchase orders. This was not yet possible with the new data cube.

For the interviewees, easy-to-use performance measurement was a clear development need from the legacy systems and was seen as an important aspect of performance measurement. On several occasions, interviewees pointed out that if the dashboard is hard to use the incentive to use it in the future decreases. Some interviewees noted that they like pie charts, bar charts, maps, and tables. Some did not want tables at all. To some, the visualization did not play a huge role. In the interviews and the workshops, it was mentioned that the filters should be easy to understand and intuitive to use. To make the performance measurement easy to use and familiar, it was decided to follow a corporate level theme on measuring, visualizing, and filtering. On the topic of visualization,

the bar chart was chosen as a dominant form of visualizing data, because of the understandability and prevalence in other organization level performance measurement. Build-in features in Power BI ensured the easy-to-use aspect of the dashboards. This included, for example, explanations to the end-user, drill-down options to explore data in more detail, filtering options, and resizing metrics.

To find out what kind of indicators would be easy to use one proposal was to divide the measures into performance measures and result indicators. The director of supply chain & QHSE (P03, Table 2) mentioned the following:

"To me, the difference between performance measure and result indicator is that based on measure's information you can derive action. The indicator tells the direction where we are going and amounts, but it is harder to derive action from them".

The director used the monthly procurement spend figure as an example of an indicator. The monthly spend figure is illustrated in the Figure 17. The chart tells the volume of purchases and trends for the selected time period, but from that, it is harder to derive action. If the spend is x amount, it does not directly tell what should be done to fix the value. OTD was seen as an example of a meter, from which action can be derived. The OTD is illustrated in the Figure 17. It is possible to drill down to OTD numbers and see which suppliers have lover OTD. Based on the finding's actions can be directed to suppliers and responsible buyers.

To make performance measurement easy to understand the case company had categorized metrics. The metrics were categorized into main metrics and supporting metrics. The supporting metrics are pieces in the puzzle and the whole puzzle is the main metric. A director of the global supply chain & category management mentioned that the CCC is a supporting metric. Savings is the main metric, which the CCC supports. Savings includes the projects that improve the competitiveness of the company. This categorization helped to support the strategic and operative procurement by illustrating which metrics have an impact on main metrics. Thus, it is easier to identify what area is causing the decline or improvement of the main metric.

As illustrated in the Table 7, the performance measurement has to be harmonized and users must have trust in the data quality. In the second business unit workshop, an issue about business unit measuring responsibilities arose. Misunderstandings in the calculation logic of some metrics led to stakeholders calculating measures differently. This has led to different results across the business units. Misunderstandings in the logic of calculations have led stakeholders to include values outside of their business unit responsibility areas. This is illustrated in the Figure 13.

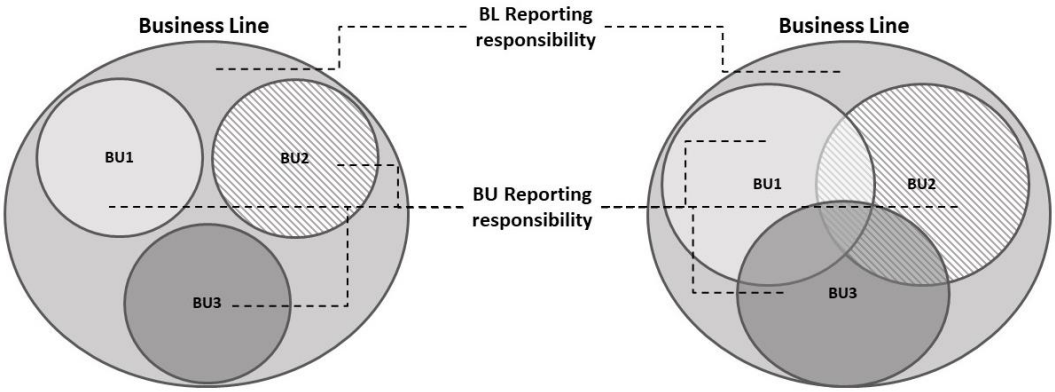

Figure 13. *Misunderstanding of reporting responsibilities.*

The misunderstanding was about the responsible areas of the BUs. For example, if a BU1 assembles products A and B in Finland. Product B consists of 100 % from CCC components and A 100 % from non-CCC components. The products cost the same amount. Therefore, its CCC percentage from the total spend would be 50 %. The BU1 decides to seek savings and decides that product B should be built in China, for the lower TCO. Now the building process of B is transferred under the company's China business unit (BU2), and with the logic of Figure 13 left side, the product is out of BU1's books. Now the BU1 is building only product A in Finland. Therefore, if one of the organization's KPIs focuses on measurement which measures CCC % from total spend, it will decrease for the BU1. The new value would be zero. In this case, the BU1 has no incentive to worsen its numbers and no motivation to seek savings from those countries. Outsourcing will have negative impacts on the business unit if the outsourced products are calculated to be outside of its responsible area. Therefore, BU1 calculated this outsourced production as their own and included it into the KPIs. This led to a situation where the same values are taken into account several times by different business units, which inflates the numbers. Two different BUs might show the same values even though the values should be shown on the BU's KPIs who is responsible for building the product. For the upper management, inflation of numbers is challenging because they have to depend on the data and make a decision based on it. This illustrates the other side of performance measurement. They can lead to unfavored actions and end up harming the company if the logic behind the calculations is not clear to the end-users and developers. By taking the same numbers into account several times it harms the requested harmony that was pointed out in the interviews and workshops. This calculation difference finding sparked important conversation in the case company to assess the issue.

4.3.1 Development

The interviews had several questions about the future development needs. They were placed to get the key users' opinions on development needs. One of the more direct questions was to ask about how the interviewees would develop the current performance measurement in their function. The question (Q6) is illustrated in the Table 3. Several development criteria rose also from the workshops and personal meetings. Usually, a personal meeting was held because the end-user had started really using the dashboards and found development areas that were missing. One of the development areas involved category management and modifying a table to be more suitable for category management. A category manager (P07, Table 2) mentioned the following:

"We should include the receipt purchase order amount because, for the supplier, the purchase order amount is not realized money".

From the supplier's point of view, the purchase order includes the canceled and ordered purchase orders. For the supplier, this value is not as important compared to the received purchase order amount, which is realized money. The purchase order can be used as a forecast for the supplier. When the category manager is in the meeting with the supplier, a receipt purchase order is looked at with the open purchase orders. The open purchase orders show the forecast for the following months. From the values, there is a possibility to find trends and abnormalities.

One development need that occurred in the interviews and the workshops were the ability to drill down to lower levels of the measurement. By this option, the end users wanted to see from which entities the upper-level values consist of. For example, the director of procurement (P04, Table 2) wanted to see from the buyer and supplier level, that how his business units spend develops. There was also a need to be able to download data to excel to be analyzed. Based on interview feedback and the workshops, although the case company has emphasized Power BI and data cube development, there was still a need for Excel or an option to calculate and drill down into the details behind the calculations. For example, in data validation, this option is valuable, where a lot of data has to be analyzed and compared with legacy systems or real-life counterparts. These needs were usually satisfied with the inbuilt features in the Power BI itself. Every indicator is possible to be downloaded into Excel, which shows the values in rows and attributes in columns. As mentioned by the category manager above, the table is useful for monitoring and presenting information to the suppliers. It is a fast way to check the open purchase order with the supplier. In the table form, it is easy to identify the possible problems. This kind of process was done successfully with the suppliers for years, and it was found to be an efficient way of communicating with other visual measures. There is also an option

to download an Excel pivot table from the dashboard. For some users, the pivot table was a familiar work environment and helped to figure out the new system and logic.

Forecasting as a part of the measurement was a frequent answer in the interviews when asked what the most important measurement in the function are. It was also frequent when asking about how one would develop current performance measurement. The forecast was seen as a heads-up for the future. It gives flexibility in operating the function when the director or manager can see trends and the number of open purchase orders. The Director of procurement (P04, Table 2) mentioned the following:

"We have long term estimates but are lacking automated short-term estimates. It is possible to gather it up manually, but it should be automated".

The director pointed out the need for automated, and thus always ready forecast to estimate the future volumes. Due to limitations in the development process, the forecast feature was not implemented in the measures. Therefore, users have to rely on manual work until the data cube has the correct features.

Harmonized performance measurement was the objective for the thesis, and it also was a clear need based on interviews. The harmonization was pointed out by several interviewees. It was requested that the management level performance measurement would be the same across the business units. The underlying data and visualization would be the same, but they would be differentiated by filters. One director saw a possibility that at the operative level there would be some differentiation in the metrics between business units. The reasoning behind this was that the business units work in different fields and thus require certain different metrics. For example, some teams worked across the business lines and needed to require those numbers as well. By harmonizing the measures the cross-comparison between different business units would become easier.

Visualization was a common topic when asked about how to develop the performance measurement in the business unit (Q6 in Table 3). The common topics that occurred were a limited number of metrics, the use of different colors, targets for the measurement, and trends. The director of supply chain & QHSE (P03 in Table 3) mentioned the need to divide the performance measures into different pages according to the categorization of the measures. The must-win KPIs would be in one dashboard and the supporting measures in other. This would reduce the volume of the measures in a dashboard, make it easier to read, and understand. To make measures even easier to read the director suggested grouping the metrics according to what they are showing. This is visible in the dashboards illustrated in the Figure 17 and Figure 18. The measures were grouped by spend and savings. The category management dashboard had an additional group

for the supplier metrics. For the end-user, the grouping makes it easier to read the dashboard and more visually pleasing.

The request for colors came from the intention to show values as red when they are below the target and green when they are above. This would pop out the problematic ones and focus more attention on the subject. This is illustrated in the Figure 14 below.

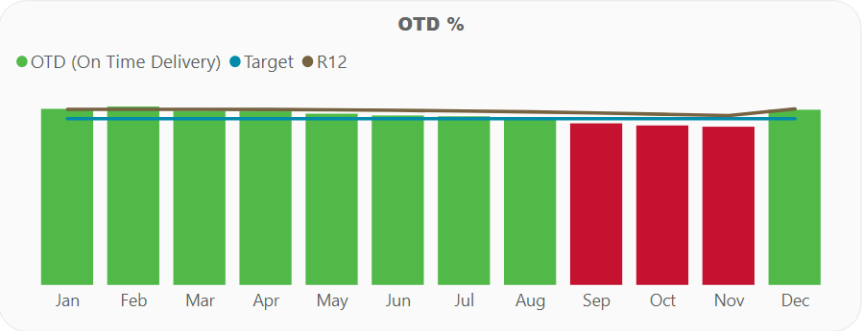

Figure 14. *Use of color to direct attention.*

As seen from the Figure 14, the months that have underperformed are highlighted clearly. From the end-user point of view, they draw a lot more attention towards the problem. It has to be noted that there could be periods where most of the months are underperforming. In this case, the whole picture would be red, and the impact of colors are reduced. This is especially the case with the targets that are reachable in the long run but not in the near future. In this case, it would not be useful to use the colors. The Figure 14 illustrates a measurement with a target and trend as well. Those features were frequent requests by the key users. The target shows a clear number of what to reach and the trend where the current 12 months are leading to. By combining these two the end user can monitor in which direction the measured values are going in the long run and at the same time the end user can see if the measure reaching the previously agreed target. The values that have trend included should be used for a longer period so that trends are visible. Above mentioned director (P03 in Table 2) noted that it would be important to use the selected metrics for a longer period so that the trends come clearly visible. For a measurement that is changed in a year, it is not feasible to measure a 12-month average.

To support operative procurement, the team managers needed to do manual work to monitor the different types of purchase orders. In the legacy system, there was no automated option to monitor these at the operative level. The senior manager (T1P07, Table 6) requested in the workshops that knowledge of the purchase order type is needed because different types of purchase orders are more time intensive. For example, system purchase orders require much more time compared to normal purchase orders. Able to

monitor this would help with workload planning. This feature was postponed because the data cube lacked the option to filter all purchase order types.

4.3.2 Risks involved

Risks of the performance measurement development process were asked in the interviews. This was done to gather heads-up information about the possible shortcomings of the development process and get valuable information about the fears that key users might have for the performance measurement.

The most frequent uncertainty was the ease of use, clarity, and user-friendliness of created dashboards. Several interviewees brought up the importance of user-friendliness for the performance measurement. The measurement has to be easy to understand, use and the dashboard should not be cluttered with metrics. Based on interview results and feedback from workshops, the high volume of indicators was seen as a negative thing. The high volume of the indicators was seen as an unnecessary distraction. The director of supply chain & QHSE (P03, Table 2) mentioned the following about the volume of the indicators:

"If too many things are measured, we are not giving the user a clear indication on what indicators are essential. The fewer indicators we have the clearer picture is conveyed to the organization about the situation in the function. Too many indicators can indicate on failure to prioritize correctly".

The fear that the director had was that the dashboard would be cluttered with new and legacy measures. The fear of the director was not false. As mentioned before, when the end users started to use the legacy system, some of them noticed missing legacy measures. Therefore, some requests for the legacy metrics were requested. In some cases, the reasoning was that it was in the legacy system as well. This finding is similar to Neely's (2000) findings in the implementation process, where managers found value to the legacy system's measures although there was no real use for them. In the research, the associated costs for operating and creating the measures were expected to be low and the value they provided high. The reasoning behind this is that there must be a reason why the measures were created and thus they must have value. In the development process, the high volume of unnecessary measures was tried to be avoided. In addition to the volume, lack of user-friendliness was seen as discouraging the use of the dashboards, making the whole development process useless. To an employee who is not familiar with the software, user-friendliness is important. This is important in a bigger company with employees from different backgrounds. The case company had this kind

of situation. Thus, the dashboard had to be user-friendly and easy to understand for users from different backgrounds.

The second most frequent risk that was illustrated by interviewees was about showing the right data and right calculations. As mentioned before, the data must be right for different business units across the business line. By failing in showing the right data and calculations, some interviewees feared that it would give the end-user the impression that the dashboards cannot be trusted. This would lead to a situation where dashboards are not used.

Need for the performance measurement to be accurate and show the right data was highlighted in the interviews. The accuracy of the data is a continuous process. In the case company, data accuracy problems occurred when a new ERP system was introduced. The new ERP had different logic built into it compared to the legacy ERP. This made it hard to harmonize the data that were gathered into the data cube. Data accuracy was also lowered by other difficulties than ERP. Due to technical difficulties with the new software or different practices, incoming dates of items were not inputted into the software on the same day. They were inputted the day before. A similar kind of situation was also observed with the direct deliveries to the sites. The day of receiving direct delivery was inputted wrongly. This was due because the practice of inputting the dates was changed. In the new practice, a new assignment had to be created and sent to the central delivery team who inputted the data into the system. This led to a situation where the wrong day is added to the system. This in turn weakens the accuracy of the performance measurement when the underlying data is wrong. For example, this affects the OTD of the suppliers, even though it was not the fault of the supplier. A director of supply chain & QHSE (P03 in Table 2) mentioned the following:

“It is important that the underlying data that is inputted to the performance measurement system is right so that the KPIs are right”.

As mentioned before, the wrong results in KPIs might have a negative impact on the success of the company. Thus, the data in the performance measurement must be right, and it has to be validated. The filtering can also impact the accuracy of the data. The Director of the supply chain & QHSE (P03 in Table 2) illustrated the need for the right filtering because it can affect the correctness of the data. It is possible to not include certain data and only include the data that improves the measurement, as illustrated in the chapter 4.3.

4.4 Implementation process

For the case company, a similar framework was introduced as presented by Parmenter (2020) and illustrated in the Figure 7. Few modifications were done to suit the implementation process of the procurement function. The refining process from the Figure 7 was removed because it would be out of the time scope of the thesis. Selling the KPI project to the owners was also removed because the owners of the project were already onboard and actively participating. Task 2.2 from Figure 7 was removed from the Figure 15 and replaced by the interviews. The task of finding the key result indicators and the performance indicators was moved after the task of designing the reporting framework. The task 3.5 was moved to the start of the third stage. This was done so that in the workshops the illustrated measures would follow a certain framework and would be consistent. The maintenance section was removed due to the scope of the thesis. As mentioned, a maintenance channel was created, and yearly review and update sessions were planned.

Stage	Task	Project month							
		pre	1	2	3	4	5	6	post
1. Getting the organization committed to the project & training the	1.1. Need for performance measurement project								
	1.2. Performance measurement lead train themselves with the help of other stakeholders								
2. Ascertain the operational critical success factors	2.1. Locate existing success factors and desired external outcomes from documentations and interviews								
3. Determining measures and making them operational	3.1. Design the reporting framework								
	3.2. Find the key result indicators (KRIs) and the performance indicators (KPIs)								
	3.3. Hold a performance measurement workshop to drop dysfunctional measures and to gather feedback								
	3.4. Refine the measures according the feedback from workshop								
	3.5. Hold a second workshop to refine the performance measures and the calculation logic								
	3.6. Facilitate appropriate use of the selected measures for end users								

Figure 15. *Process for performance measurement development* (adapted from Parmenter, 2020)

As illustrated in the Figure 15 above, the implementation process was divided into three sections: Company commitment, critical success factors, and implementation process. In the first section, meetings were held to determine the procurement need for the KPI project. A major part of this was already determined before the thesis. The second task

of stage one in the Figure 15 was done before and during the interviews. Stakeholders' previous knowledge of performance measurement played a key role in the process. The training of the personnel happened during the first months of the project.

The second section in the Figure 15 was about deciding the critical success factors and holding the interviews. Qualitative primary data were gathered through interviews. In the interview, feedback was gathered from the company's current situation, preferred development paths, preferred metrics, measurable things, and possible risks involved in the development process. The performance measurement design process framework of Neely (2000) was followed when deciding the themes of the questions. The questions are illustrated in the Table 3. The feedback gathered from these questions exceeded the expectations and showed the involvement of the organization in the project. The interviewees provided previously used performance measures. This included legacy systems manual calculations and other presentations/statistics that the interviewees were reporting. Materials were analyzed with the interviewees. Analyzed materials gave a solid understanding of the reporting needs of the interviewees and thus gave a good foundation for the next section.

The final section was about the implementation process itself. In the first task, the reporting framework was designed. Design frameworks from the literature were used. Neely's (2000) performance measurement system design process was used for the design of the performance measurement. Parmenter's (2020) related levels of performance measurement are used to define the number of metrics. This is illustrated in the Figure 8. The KPIs and KRIs are identified based on Parmenter's (2020) definition of performance indicators and result indicators, visible in Figure 5. For the visualization, the visualization framework by Abela (2008) was used. Based on the frameworks and feedback from the interviews, draft dashboards were created for the workshops. In the workshops, draft dashboard metrics were assessed. Between the workshops, the draft dashboards were developed based on the feedback. In the final workshops, final feedback was gathered, and the usability of the dashboards was illustrated. As mentioned, the task of facilitating appropriate use was partially done in the final workshops and the private meetings. Additionally, a document was created that illustrates the functionality and calculation logic behind each metric. This document is illustrated in the appendix.

4.5 Creation of dashboards

The Neely's et al. (2000) framework was used as a guideline for the design process, illustrated in the Table 1. From the framework, one of the major requirements was that the measure had to be connected with the organization's strategies and objectives. A

clear purpose for each measurement was required. With this approach, the volume of the measures could be controlled. The legacy system had a heavy focus on financial measures, therefore measures based on ratios and non-financial measures were tried to include in the dashboards. From the viewpoint of maintenance, the focus was to make the dashboards as easy to maintain. This would ease up the revisits and modification of the dashboards in the future. The ease of use was also a key driver behind the metric design. The ease of use was the main driver when selecting filters and the overall layout of the dashboards. Visualization was considered and the best visualization was selected based on Abela's (2008) framework.

Based on interviews, a total of three draft dashboards were created. As mentioned in the previous chapter, the draft dashboards were used as a platform from which the dashboards can be improved in the workshops. It was not intended to cover all aspects that were presented by the interviewees. Several interviewees pointed out that the most important metrics are the CCC-%, Annual spend, savings, and OTD, as those were part of the organizational strategy and must-wins. Other lesser requested ones were the monitoring of VCS (Case company's category selection) coverage, COGS, savings % by spend, spent by region, top 15 suppliers by spend, and payment terms. All of these metrics, except payment terms, were incorporated into the draft dashboard illustrated in the Figure 17. Due to the lack of feature in the data cube the payment term measure was not included. The requested metrics varied between the different dashboards. For example, the category manager wanted to see spend according to the categories and subcategories. For the operational procurement, there was a need for a table where all the team's buyers and suppliers would be visible was seen as a good tool to use in team management. Management would see spend by a team with the open purchase orders. An observation was made that most of the requested metrics were financial metrics (CCC, spend, and savings) but metrics based on ratios were still present (savings % by spend and OTD).

The dashboards were iteratively developed first at the workshops and later, based on private meetings and feedback. For the feedback, a development channel was created for easier communication. This channel also worked as a maintenance platform for the owners of the dashboards, where solutions would be shared and requested updates assessed. Some of the requested metrics were not able to be implemented due to a lack of data or software capabilities. Illustrated dashboards in the chapter are created in a demo environment. No real values were used to protect the privacy of the case company.

Three subpages were created per dashboard, except for the team dashboard. The data cube lacked critical features so that the subpage for receipt purchase order would be

possible. The subpages are illustrated in the Figure 16. Idea was to create one subpage for calculating the purchase orders, the second with the receipt purchase orders, and the third with spend (invoices). This was done to help with the validation process and to harmonize the business unit level performance measurement with the organization level performance measurement. Additionally, receipt purchase order and purchase order subpages allowed familiar values to the end users, which was intended to help the launch of the dashboards.

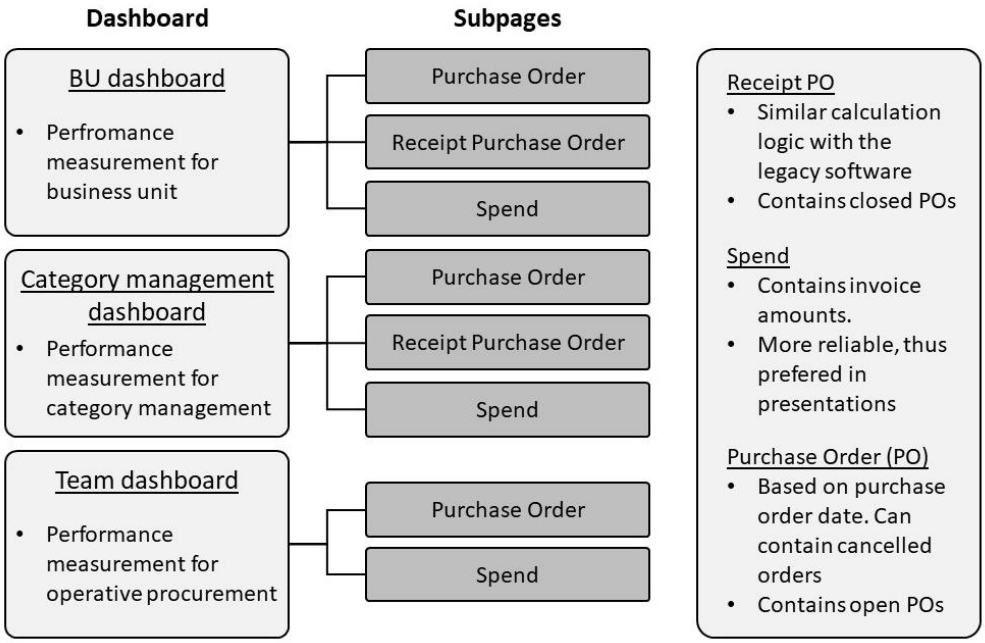

Figure 16. *Developed dashboards and subpages.*

All the subpages within the same dashboard have identical metrics in them except the team dashboard. Due to technical difficulties in connecting the teams with the right receipt purchase orders led to the removal of the receipt purchase order subpage for now. When the technical difficulties are removed the subpage will be implemented. In the Figure 16, the subpages with the receipt purchase orders are in line with the legacy system numbers and therefore the data validation is easier, and the numbers are in line with the historical data. As illustrated in the Figure 16, the receipt purchase orders contain only closed purchase orders. This means that there are no values that are not realized in the figures. This increases the reliability of the values. The purchase order subpage was implemented to harmonize the performance measurement with the organization level performance measurement. Purchase order calculation logic is used widely in the organization level performance measurement. As mentioned in the chapter 4.3, one downside that the purchase order has is that the values contain open and sometimes canceled values. This inflates the values, although the magnitude of the effect is small. The advantage that the purchase order has is that includes open purchase orders, which can

be used for forecasting. The spend is used for presenting the performance measurement for the stakeholders because it contains only the invoiced amounts. Therefore, it is more reliable, although it has a slight delay compared to the receipt purchase order. This is because the invoices are usually received later than the purchase orders. Overall, the values that the spend subpage illustrates are close to the receipt purchase order sub-page values.

When designing the business unit dashboard, a major part of the results indicators was gathered under the CCC -group title. This section was intended to give the end-user an understanding of the volume and trend of the BU's procurement activity. The spend money was also divided into the CCC and non-CCC. Thus, the end-user can have additional information on money spent on cost-competitive countries without sacrificing the overall picture. Users can drill down from year figures down to monthly figures. The second group for the results indicators was the "Savings" -group. Savings group and the CCC group are illustrated in the Figure 17. Savings illustrates savings made by the BU on percentages and per month. Savings are further divided into categories where the savings came from. These categories are pooling, project logistics, and spot prices. Monthly target saving is visualized in the monthly savings chart. The target changes according to the selected business unit and selected timeline. The last result indicator, VCS coverage was placed at the bottom of the dashboard.

One of the important performance indicators OTD was placed to the center location. Users can drill down from year level to monthly level and see the target and rolling 12 months. For supplier performance monitoring, a metric of the top 15 suppliers for the business unit based on spend, was placed into the dashboard. With the metric, the user can see who the biggest suppliers to the business unit are and by how much. The metric helps to identify unfavorable trends among the suppliers.

The first dashboard had clear filters for the business unit, business line, timeline, legal company (which region), and supplier selection. These filters are visible on the top of the dashboard, in the Figure 17. These managed wide control of data to the end-user. For easy-to-use purposes, buttons for business unit selection were integrated into the dashboard. These buttons select the business units automatically when pressed. Similarly, buttons for viewing data on periods (selected timeline), year-to-date (YTD), and rolling 12 months. Each metric would update according to the selection. The illustrated draft dashboard in the Figure 17 has the YTD -view selected.

The balance between manual and automated measurement was mentioned before in the interviews by the director of the supply chain (P02, Table 2). Similar balancing consideration was done in the implementation process. There were several requests to automate the manual process, most notably the savings process and the managing process of business units. The former was tricky because the information of the savings was gathered from several managers, category managers, and directors to one Excel. This made it hard to get to the end source of the savings and automate it completely. Thus, a decision was made to refer to this one Excel, which would be manually updated by the users. In that case, the dashboard would show the result of that Excel to a wider audience and make it more visually pleasing to present. By doing it this way the maintenance of the dashboards would be easier even for a less skilled person. This enables easier revisit to the dashboard. The successful automation process was to automate the business unit monitoring. The example mentioned in the chapter “Implementation process” by the supply chain director, mentioned the manual work to create the KPIs. The process was able to be streamlined and the manual process was eliminated completely.

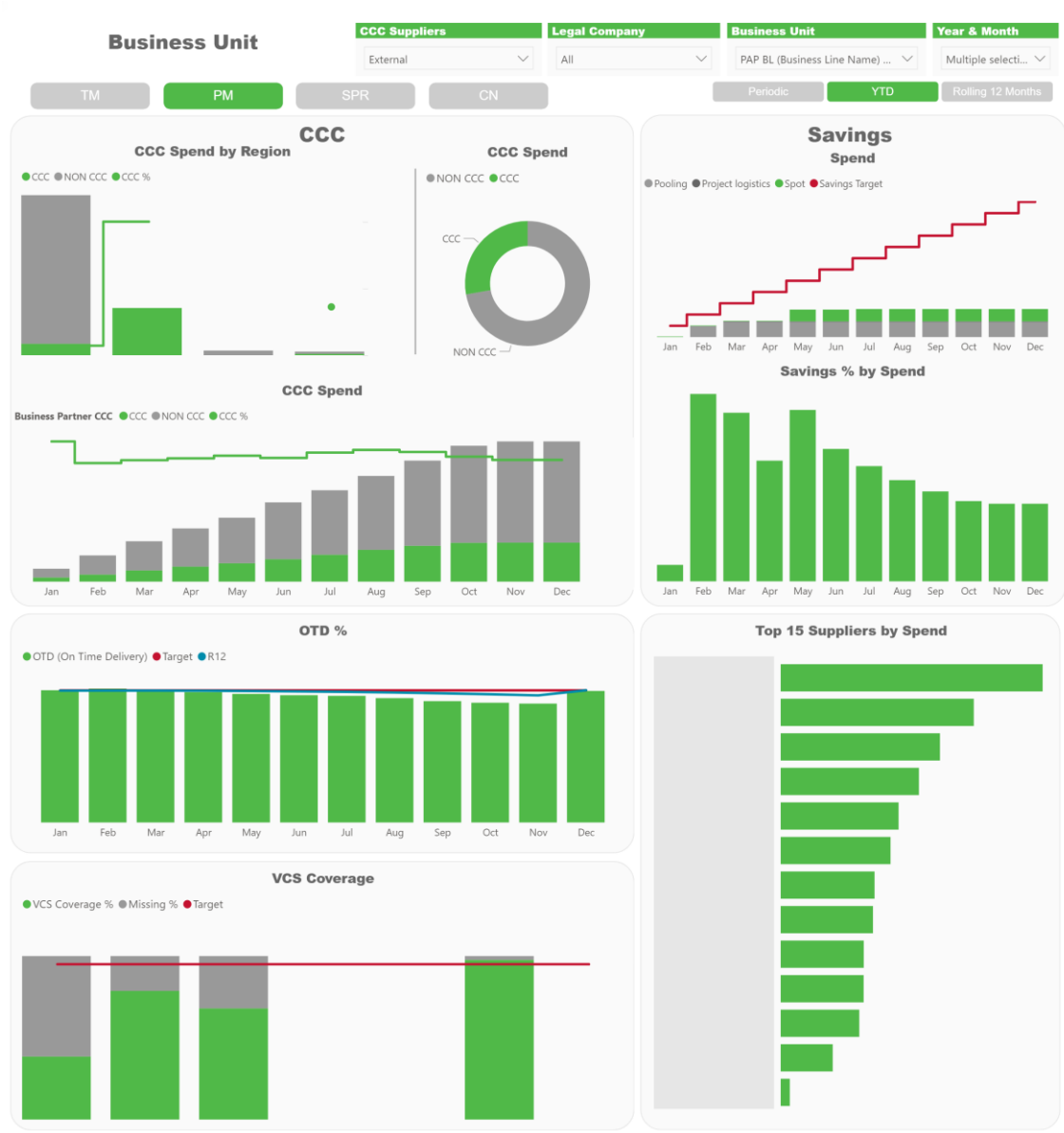

Figure 17. *First draft dashboard based on interview results.*

Between the first and the second workshops, the requested features were analyzed and added. The duration was from one to two weeks depending on the dashboard. There were several changes from the draft dashboard (Figure 17) to the version after the workshops (Figure 18). Most notably, additional measures have been added and the VCS coverage metric was removed. In the workshops, it was noted that there is no longer a need to monitor VCS coverage on the BU-level dashboard, due to the high rate of coverage. This was in line with the Table 1 framework of requiring purpose for the measures. The additional metrics were supplier classification metrics and a table for more detailed monitoring of the functions under the BU. Based on the feedback of the workshops some visual changes were made to the OTD metric and Top 15 Suppliers by Spend metric.

The change to the OTD metric was done based on Abela's (2008) visualization framework. In the framework, a line chart was preferred if the values are changing over time and there are many periods illustrated. The target for OTD was removed with intention of reducing maintenance due to the need of changing it over time. Additionally, the driver was that the upper management reported the OTD values without a target. The removal of changing targets also affected the use of color labeling. The use of color-coding based on targets was discussed based on the feedback from the interviews. It was decided to discard the color coding to lower the maintenance required for the dashboards. In the case of the OTD, the set target is usually high and therefore the majority of the bars would be red. This would defeat the purpose of the color-coding as mentioned in chapter 4.3.1. To increase the accuracy of the data, different types of savings (pooling, project logistics, and spot) were removed. The definition of them was not accurate and thus it was seen as better to remove them. Additional filters and buttons were added to help user-friendliness and filtering of data. For example, the added button "PAP" automatically selects the whole BU data. Titles were modified to suit better the presentation of data to other stakeholders. This is illustrated in the Figure 18, where the pie chart's old title of "CCC Spend" was changed to "Total Spend". Additionally, one filter was added to the top of the dashboard. This filter allowed the selection of different teams in the procurement. More filters were requested in the last workshop, but they were included under the advance filtering option, which the user can open if needed so. By this approach, the dashboard could remain easy to use and understand to a wide range of users.

A histogram was requested in the workshop of the business unit dashboard for monitoring the distribution of suppliers in the business units. For seeing the distribution, a column histogram chart was introduced based on Abela's (2008) framework in the Figure 9. Based on the framework, a comparison between top suppliers was done with horizontal bar charts. The composition of spend was done with a simple pie chart. Although a lot of measures are in line with the framework, the dashboards have differences from the framework. One of the most notable ones is the excessive use of the bar charts. The use of bar charts was justified by the closeness of the periods. Therefore, the trend is still clearly visible in the bar charts. With the bar charts, additional information can be presented in the same chart. In the Figure 17 and Figure 18 additional information was to show the CCC amount and percentages. Additional, the reasoning behind choosing the bar charts was the corporation theme style of using the bar charts over the line charts. Aligning the business unit performance measurement with the corporate theme style ensured the harmonization between different dashboards and ease of use. This was preferred with the cost of line charts.

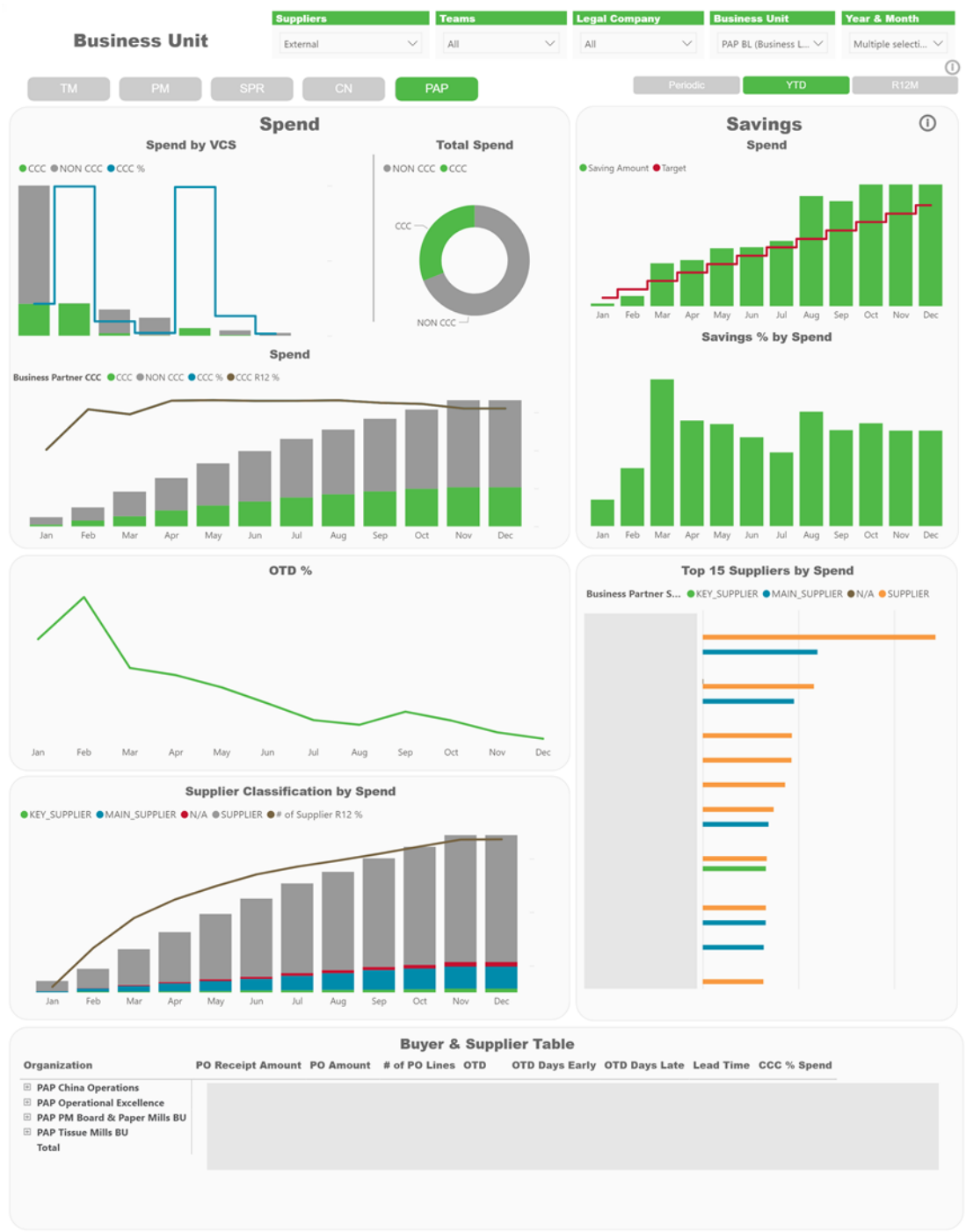

Figure 18. *Dashboard after the workshop meetings.*

As can be noted from the updated features, the balance between usability and maintenance was one of the key factors. The aim was to avoid the situation where the dashboard would be outdated, because of difficult maintenance. In the workshops, outdated performance measurement was seen negatively impacting the utilization rate. A director (B1P07, Table 4) mentioned that too many development projects have failed due to too hard maintenance requirements.

The other key factor in the development process was to find the right balance with the number of indicators. If the number of indicators was not regulated the dashboard would be full of metrics and the focus would be off. To regulate the number of metrics in the dashboards, Parmenter's (2020) framework on different levels of performance measurement was used. As illustrated in the Figure 8, the business unit dashboard should have under 15 performance measures. In the development process, some metrics got discarded not only because of their challenges but because of their purpose. For example, the VCS Coverage metric was removed because the BU dashboard was not the right place for it.

One of the lacking features in the Figure 18 is that the metrics do not have a forecast in them. Although the illustrated dashboards are based on incoming invoices and thus forecasting is more difficult compared to purchase orders. In the case company, the forecasting has been done by open purchase orders. The lack of forecast is due to technical difficulties in the data cube development. This led to delays in implementing the open purchase order into the dataset. Another lacking feature due to the limitation in the data cube was the average payment term measure. The indicator is one of the KPIs for the business line management and therefore it would be part of the business unit dashboard.

Each key user involved in the workshops was involved in the data validation process. This was done due to the sear size of the data and because the key users were already familiar with the historical number. Thus, this made the validation process faster. Data was also validated from the general levels and occurring differences were investigated.

5. CONCLUSION

Conclusion summarizes the key findings of the development process in the case company. These findings are analyzed and compared to the literature. Lastly, criticism and limitations of the research are presented with the possible future development paths for the research and the developed dashboards.

5.1 Key findings

In the development of the performance measurement, a corporate strategy played a key part in what measures got included in the performance measurement dashboards. All the dashboards had key measures that were tied to the key corporate strategies and objectives. These were the CCC measures and OTD. But not all performance measures were defined by the corporate strategy. A couple of measures at the business unit level were defined by the need of the end users in the procurement function. For the category management and the team management, the performance measurement included more measures that were derived from the operative needs of the function. The aim of this was to align the objectives from strategic procurement to operative procurement but also include end user requests into the dashboards. Thus, the main objective of the thesis was:

"Create harmonized and coherent procurement performance measurement dashboards to reduce manual work and to align strategies and objectives of operative and strategic procurement".

Due to the wide range of procurement functions in the different regions around the world the case company had an incentive to harmonize the procurement performance measurement. Harmonization would allow cross-comparison of numbers, coherent calculations behind the measures, and therefore more accurate results. The harmonization process was put forward by the case company by the creation of a data cube, which gathered data from different sources under one system. This data source was used as a dataset for the thesis performance measurement and around the company. The underlying dataset ensured that the data is similar between the dashboards, the difference being filters and additional calculations. For example, the organization level dashboards used the same dataset. Therefore, harmonization towards the organization was ensured. The strategies and objectives of the organization were integrated into most measures in the dashboards. The remaining measures were based on the requests of the end users.

This approach formed a redline line from organization-level measures to operational-level measures. This was in line with the framework illustrated in the Table 1, where the corporate strategies and objectives were seen as a key driver for performance measurement. To ensure coherent performance measurement, the created dashboards followed the organizational theme which was already implemented at the organizational level. This approach ensured that the visualization and functionality remained the same between the levels. The aim of the objective was also to reduce the amount of manual work needed on performance measurement. As mentioned in the chapter 4.1 where the current state of the performance measurement was analyzed, the director (P02, Table 2) spend hours building the measures each month. This was reduced substantially after the implementation of the dashboards. Therefore, the first objective of the thesis was reached.

There are many aspects to developing a performance measurement system. In the literature review, several authors pointed out that the performance measurement should be connected to the corporate strategies and objectives (Bourne et al., 2000; A. Neely et al., 2000; Vitale & Mavrinac, 1995; Weele, 2018). By connecting the corporate strategies and objectives to the performance measures, the success of those critical factors can be monitored and managed. This ensures that the focus is on measures that matter (Parmenter, 2020). The same need arose from the interview feedback. Several directors pointed out a need for corporate strategies to be visible on every level of the performance measurement. The main research question for the thesis was:

“How to develop a procurement performance measurement system in a company operating in a manufacturing project business”?

The illustrated frameworks in the research background guided the process of developing the performance measurement system for the case company. The performance measurement design process framework, illustrated in the Table 1, gave desirable features for the performance measurement system. The features were used when designing the dashboards for the workshops. The performance measurement implementation process framework was modified to suit the case company and the business line in hand. The modified version of the framework is illustrated in the Figure 15. Modifications were done to the first part of the process. This was done because the management was already onboard with the project and therefore there was no need for the first steps. Additionally, some tasks were rearranged. In the development process, clear tasks were found to be useful for guiding the process. The framework was seen as a useful tool, but it requires modification to suit the different situations. Although modifications are sometimes needed, the core structure of the framework should be maintained. These core structures

are the commitment of the organization, ascertaining the critical success factors, determining the measures, and making them operational. Within the last stage, it was found useful to hold an interview, which would be the foundation for the dashboards. These dashboards would then go through two workshops, which include key-users from the functions that are going to be measured. Therefore, the end-users are involved, and they can say their opinion and propose additional measures. There also should be a second channel (personal meetings, messaging channel) to propose measures and ask for help. The workshops would also illustrate to the users how the dashboards should be used and what are the calculation logics behind the measures.

To answer the main research question comprehensively, three sub-questions were created. The first sub-research question SQ1 was answered based on the gathered qualitative primary data from the semi-structured interviews and personal meetings. For the interviews, key users were selected from key procurement functions. To be able to develop the procurement performance measurement the current situation of the performance measurement in the case company has to be known. Therefore. the first sub-question for the thesis was the following:

“What is the current situation of the procurement performance measurement in the case company”?

The case company has a legacy performance measurement system in place. Key users noted in the interviews that the coverage of the legacy system for procurement is good. Key users found only some additional measurements that were missing from the performance measurements. The one that was mentioned frequently was savings and accurate CCC -measures. The main problems that occurred from the current situation were the lack of harmonization and required manual work for building the performance measures. The legacy performance measurement system lacked harmonization between procurement functions. This was due because in the legacy system the end-user was expected to filter data and partially build the dashboards. This situation created lots of different measurement logics and metrics. This lack of harmonization made cross-comparison difficult and in some situations gave wrong values out of the measures. The second problem that plagued the legacy system was the manual work required to build performance measures. The different legacy ERPs did not communicate together and thus required in some cases the lengthy process of manual work to create the wanted measures. Data had to be filtered and added, which increased the probability of errors. One director used several hours at the start of each month to create different KPIs for reporting.

In sub-chapter 2.8 it was mentioned in the literature that the corporation might be in the middle of software and database updates, which might reduce the accessibility of the data. This was the situation in the case company. The case company was in the middle of data warehouse development and new ERP system implementation. This created some issues in the development process when the old legacy ERPs were used, and they did not communicate with the data cube. The issues lead to reduced accuracy in some regional areas, excluding some regions totally. The data cube was in the development phase at the time and did not have complete coverage of the procurement metrics. However, the impact of this was mitigated by inputting manual data from users. This was done as a temporary solution until the data warehouse was complete enough. The second sub-question was about the need for the performance measurement:

"Why the internal procurement stakeholders need performance measurement"?

According to Komatina et al. (2019), the fundamental objective of performance measurement is to indicate areas that do not reach the target values. Areas that are reaching their targets can be illustrated for bonuses. In addition to showing areas that are underperforming, according to the literature and the feedback from the interviews and workshops, the performance measurement is needed for reviewing strategy implementation and increasing visibility and harmonization across the procurement functions (Bourne et al., 2000, p. 758; Vitale & Mavrinac, 1995). Vitale and Mavrinac (1995) mention that the performance measurement glues strategic objectives into functional tasks by redirecting attention into critical outputs necessary for the strategy. The need to monitor the strategy implementation was mentioned as a key need for the performance measurement.

The need for the performance measurement was not only seen as a reporting tool to the upper management but as a platform for conveying procurement strategies from top to bottom. By conveying the strategies and objectives from the top to the bottom the objectives and strategies of strategic and operative procurement could be aligned. As a director (P02, Table 2) mentioned: *"We have this alignment or redline from business line top to bottom. So, we are working with the same priorities".* For the objectives and strategies to be trustful a harmonization is needed to avoid misunderstanding in the performance measurement system. Due to a lack of harmonization, misunderstandings were observed in the case company. Business units had different calculation logics behind the measured which led to differences in received values. Therefore, to ensure that the measures are harmonized a performance measurement is needed.

Visibility of the measures was one of the key aspects mentioned in the interviews regardless of the background of the interviewees. In the interviews, the director saw visibility

as a key feature for conveying strategy. The team management and buyers wanted to see how they and their teams are performing and how it affected the upper-level measures. There is no need for an operational person to see upper-level measures all the time, but there needs to be access to do so. Thus, the user can see the impact of their work.

The motivational aspect of the performance measurement was observed in the interviews, workshops, and literature. Parmenter (2020) mentions that performance measures can indicate areas that are above or below their targets and therefore be used as a motivator. This tends to motivate people into action and help people see the progress and trends (Parmenter, 2020). Similar results were observed from the vice-president's (P01, Table 2) feedback on savings. The vice-president mentioned that the purpose of the savings is to motivate the procurement to perform better. By measuring savings, the procurement function can be sure that the commodities and services are bought at the lowest TCOs or best value possible.

To conclude, the evidence from the literature and case study illustrates that the performance measurement is needed for conveying strategy and objectives, monitoring strategy implementation, harmonization, and increasing visibility across the procurement functions. Performance measurements focus attention on the areas that are underperforming. There is also a need to motivate the end users with the performance measurement. The third research sub-question for the thesis was:

"What kind of performance measures does the case company need to support strategic and operational procurement? How they should be developed"?

In the literature, there is a common understanding of evaluation criteria for performance measurement. The framework introduced in the Table 1 was used for defining these criteria. The framework noted that, the performance measurement is unique to each organization. Therefore, measures should be derived from the organization's strategy, objectives, mission, and vision. Similar observation was made when analyzing the interview feedback. It was seen as an important aspect for conveying the strategies and objectives to the lower levels of the procurement. Additionally, the modified framework suggested that there should be a balance of non-financial, financial, effectiveness, efficiency, internal and external measures. Similar aspects were observed throughout the literature. (Malina & Selto, 2004b; A. Neely et al., 2000; van Looy & Shafagatova, 2016) Although the balance of different measures was included in the modified framework, the major number of the measures were financial in the final dashboards. This tendency to prefer financial measures was observed in the scientific publications. Observation was that the

larger companies prefer the financial measures over the non-financial ones (Malina & Selto, 2004a). One future development path for the case company's measures would be to prefer the non-financial measures. The modified design framework was seen as an effective way of developing performance measures for the case company. It took into consideration a wide range of features that performance measures should have.

The above-mentioned characteristics are observable in the created performance measurement dashboards. The organizational strategies are aligned with measures (CCC, Savings, OTD) that conduct the organizational strategies. This strategy is brought to the category management and team management dashboards with options to drill down to specific values. This satisfies the first criteria of aligning measures with organizational strategies, missions, and vision. The second criteria of balance between different measure types are fulfilled partially in the developed dashboard illustrated in the Figure 18. Percentages are used to illustrate for example on-time delivery, financial values are presented in the savings and the procurement spend charts. Effectiveness and external measures are utilized in the OTD measure and the table, internal measures are used in savings measures. The one aspect that is not directly used is efficiency. One could argue that with the help of the dashboard the efficiency of the procurement can be increased. For example, by monitoring the volume of purchase orders per team.

To support the strategic and operative procurement, the right measures are needed to be used in the right situation. The common perception is that all the measures are KPIs. The literature divides measures into two categories according to how "wide" the indicator is measuring (Parmenter, 2020). These two categories are result indicators and performance indicators. The performance indicator can be tied to the specific team or teams. From them, it is possible to identify the problem. Result indicator shows values that are result from several groups. It gives an overall picture of the specific area whereas performance indicators give a more direct one. (Parmenter, 2020) The director (P03, Table 2) divided the measures with similar logic. For the director, one can derive actions from performance indicators whereas results indicators give the direction where the values are going. Director used a monthly purchase spend as a reference to a results indicator. It is much harder to derive actions from it.

When the right measures are in place, there still needs to be control on the volume of the measures. It was observed in the interviews when asking about what measures are useless that there rarely was one. This finding was in line with Neely's (2000) observations that managers found the value of current measures to be high and the cost associated with them to be low. Interviewees in Neely's research found explanations for every measure and did not value them useless. The reasoning behind this logic was that there

must be some value behind the measures because they were created in the first place. A similar situation was observed in the thesis interviews where interviewees found almost no useless measures when asked about. One could argue that the question (Q7, Table 3) was too direct and emphasized only useless measures. Maybe the question should have asked what measures the person does not use or what measures are outdated. This approach could have yielded better results. Nevertheless, it is interesting that the organization does not have useless measures and all the measures have a clear purpose. Based on the literature, it is crucial to determine the explicit purpose for each measure, as the Neely's et al. (2000) framework emphasizes. Only those measures should be included that have a clear purpose and are fit for the dashboard. Additionally, the purpose of the measures in the dashboards should be checked yearly. With this approach, the cluttering of the dashboard can be avoided, and the outdated measures would be cleared. The high volume of metrics has a negative effect on how the strategy on objectives is conveyed to the different levels of the organization. The director (P03, Table 2) mentioned that by limiting the number of metrics in a dashboard the management sends a clear message to the lower levels of the organization, about the targets and objectives. Failure to do so would convey a message that the targets and objectives are not clear. The used Parmenter's (2020) framework for the number of metrics, conveys the same message.

For the case company to have good measures to support the strategic and operative performance, the calculation logic has to be clear to everyone. The clear calculation logic was emphasized in the performance measurement system design process. In the second workshop, misunderstandings between calculation logic were found. This misunderstanding had led to a situation where values were included two times in the different business units' KPIs. The inflation of numbers is problematic for the decision-making process. It may lead to situations where decisions are made based on wrong values. The discovery of misunderstanding in the calculation logic sparked conversation in the case company to fix this situation. The finding re-enforces the need for clear calculation logic for measures.

From the start of the development process, ease of use was one key driver for the measures. In the interviews, it was mentioned several times. The interviewees saw poor user friendliness as a negative driver for the use of the dashboards. Although the ease of use was a key driver, there had to be an option for deeper analysis of the data. Some end users wanted to analyze the data on Excel or drill deeper into the data. The used software had an intuitive solution for these situations while preserving the ease of use.

Visualization was one of the key areas for reducing cognitive calculation and increasing the ease of use. In the literature, visualization has an important role in replacing cognitive calculations with better visuals while improving understandability, memory, and decision making. This is accomplished by the human ability to visualize patterns, trends, and outliers. (Heer et al., 2010) In the interview and workshops, improvements in visualization were seen as a key driver for improving ease of use. In the development phase pie charts, bar charts, maps, and tables were seen as a good visualization method for the dashboards. When deciding the visualization Abela's (2008) framework was followed. According to the framework, the proposed charts are good for comparison, composition, and distribution. Although a lot of measures are in line with the literature, the dashboards have differences from the literature. One of the most notable ones is the excessive use of the bar charts. One of the reasons for using bar charts was that the illustrated periods are close and therefore the trend was visible to the user. The second aspect that led to extensive use of bar charts was that it was the main visualization used in organizational levels of performance measurement. By following the same design theme, the harmonization could be maintained. During the implementation process, it was noted that some interviewees did not like the bar charts. Due to the harmonization effort, these proposals were discarded. Even though some differences were observed between the visualization framework and the created dashboards, it was seen as an intuitive and effective framework. One can quickly recognize the required measures for different situations.

The use of colors was proposed in the interviews. Colors were requested because they would direct attention to the critical aspects. In the literature, there are mixed results based on the use of colors in performance measurement. Based on Abela's (2008) study, colors should be only used to highlight areas of interest. In the case company, the high targets to certain measures occurred problematic. Certain measures would be most of the time on the negative and thus the use of color would be obsolete. The other concern was the maintenance issue. The coloring would have to be updated with the target and that was not seen as favorable. Taking into consideration results from the literature and the drawbacks of the coloring on maintenance and visualization, the upside of coloring was not feasible. In the future, the implementation of colors would be one development path.

When the development of the dashboards began, the balance between maintenance and features had to be considered. As mentioned in the previous paragraph, some features had to be removed due to high maintenance costs. The idea was to keep the dashboard as simple as possible so that it would be easy to maintain and update throughout

the years. A director noted that too many projects fail because they are too hard to maintain. To avoid this in future development processes, maintenance should be one aspect in the frameworks, as it is in the Table 1.

To conclude the third sub-question, the measurement should be derived from the organization's strategy, objectives, mission, and vision. There should be a mix of non-financial, financial, effectiveness, efficiency, internal and external measures. To support strategic and operative procurement, definitions of metrics should be clear. Differences in calculation logic create confusion and might mislead the decision process. Additional concern is the number of metrics in the dashboards. By limiting the number of metrics, the performance measurement dashboard sends a clear message on what are the targets and objectives of the organization. In addition to the strategic and objective aspects of the measures, visualization has to be considered. By enabling effective visualization for the right situation, cognitive calculations can be replaced. Therefore, the metrics are easier to read, easier to use, and outliers are easier to identify. The use of color was found to have mixed results in the literature. A measure that is not always below the target, it was seen as a positive to color code the underperforming periods. This would draw the attention of the end user. But if the periods were always below the target, the coloring was seen as distracting. Finally, when designing measures, the balance between maintenance and additional features should be considered. Too many features make the measure hard to understand and read, but also hard to maintain. By making hard to maintain dashboards, they are harder to update over the years and may become useless.

5.2 Managerial implications

The management should consider continuing the implementation of performance measurement systems into the procurement function. In the literature, it has been proven that by using a performance measurement system in the procurement function, the performance of the procurement increases (Bello & Gilliland, 1997). Therefore, continuing the implementation of the performance measurement system into the procurement is justified. This implementation process serves as a valuable reference point for similar future projects.

In the development process itself, the management should include the strategy of the organization and the procurement function into the performance measurement system. Therefore, the strategy is conveyed to the lower levels, and the different hierarchies of the procurement and the organization are aligned. It is also important to involve the end-users in the development process so that they are familiar with the system and most importantly feel that they have been heard. By this approach, the management can be

sure that the needs of the end-users are fulfilled. By involving the end-users, the performance measurement has a higher chance to be used.

In the implementation process itself, it was observed that tradeoffs have to be considered when designing the performance measurement system. Tradeoffs between manual and automated reporting had to be considered and how much automation is worth the effort. In addition to this, tradeoffs between features and maintenance had to be made. It is very easy to just add features and forgot the upkeep part. By including countless features, the system becomes harder to update and understand. This may lead to the abandonment of the developed performance measurement dashboards. Managers should be aware of these tradeoffs so that the developed system exists and functions for years.

Differences in calculation logic were observed in the implementation process. It occurred that different functions calculate the metrics differently from each other. Usually, there was some motivation behind the calculation logic differences. For the management, this remains one of the key issues to be solved. It is important that the calculation logic is standardized between functions so that everyone agrees and knows how the performance measurement system functions.

Short-term managerial implications for the created performance measurement system would be the maintenance of the system. As illustrated in the literature by Neely (2000) and Bourne (2000) the maintenance and yearly review of metrics are necessary for the success of the performance measurement implementation process. The relevancy of the measures should be checked yearly. The intention of this is to delete measures that serve no purpose. This kind of situation may occur when the strategy of the company or procurement changes or the focus shifts. By being one of the fundamental drivers of the performance measurement system, it is crucial that the measures convey the current strategy and message (Bourne et al., 2000). The targets of the measures need also yearly updates and validity checks. The targets move over the years and for usability, it is crucial that the targets are on point because they drive the motivation and action. For example, in the implementation process of the thesis performance measurement dashboards, some targets were deleted due to high maintenance.

Long-term managerial implications for the performance measurement system should include aspects that were out of the scope of this thesis. Those are quality and supplier evaluation. Quality was brought up in the interviews especially by the category managers. It was seen as a crucial aspect when evaluating suppliers. The decision was made to sidetrack quality over other performance measures because the implementation of the data on quality was still under development. In the case company, there was data on

quality on legacy ERP and upper corporation level. This satisfied the need. One of the future development paths would be to include quality as an aspect of the performance measurement. The further development of the dashboards should also include some measures that were excluded from the thesis project. Additional measures for managing the categories were especially requested. These included purchase agreements for seeing which items rotate and for seeing the agreement coverage. Additionally, price increases and decreases for certain products were seen as a good measure. The idea for implementing the prices was that it would lead to an increase in design-to-cost cases and an increase in savings actions. Currently, these features were not feasible to create due to technical limitations, but in the long term, these features could be revisited and evaluated.

The status of the IT systems should be taken into consideration before the development process. Are the IT systems in that state that they can manage a performance measurement system? For the developed dashboards, one of the major limiting factors was the IT problems. In the literature, those were also noted as one limiting factor to the implementation process. The IT problems stalled the development after the workshops because the dataset could only be updated by the development team. Due to the heavy workload of the development team features were added slowly. Therefore, the dataset lacked features that made the requested metrics hard to implement. One feature that was requested, but not possible to be implemented was the forecasts for the measures. Workarounds could be created but they had their limitations. By giving the dataset more time to mature, the before mentioned features could become easier to develop.

5.3 Criticism and limitations

It should be noted that no single case study can give a comprehensive picture of how to develop a procurement performance measurement system. The aim of this thesis was to implement a performance measurement system in a specific environment at a specific time. Therefore, this study is hard to replicate because of the special environment, interviews, and workshops. Thus, this study cannot be generalized to every company. However, this study can be used as a reference for companies or business units in a similar situation and environment.

To increase reliability (repeatability of the process) data was gathered from each business unit to have a wider sample size from different backgrounds and business units (Lewis et al., 2019, pp. 213–214). For example, directors from each business unit were interviewed to gather feedback from different backgrounds and thus increase reliability. A total of 24 key users from the business line’s all business units participated through

interviews and workshops. A couple more people were involved through personal meetings. The numbers fall short of the sample size of 30 proposed by Saunders et al. (2019, p. 300). As mentioned by Saunders small sample size does not likely represent the population mean as accurately as a large sample size does. Even though the sample size was not over the proposed one, it was as big as it could be. For example, in the category management dashboard development project all the category managers participated in the workshops. The reliability of the results is influenced by the interview results and feedback gathered from workshops. Interviewees have their own personal opinions and views which might lead to subjectivity. The interpretation of these results also influences the reliability. Reliability is also affected by the literature and data sources used in the thesis. To mitigate this issue the goal was to choose only pre-revied articles and sources. In addition to this high number of sources was preferred. The reliability of the used literature was also tested in the implementation process itself. This increases the reliability compared to the situation where no actual implementation process was done.

To increase objectivity the interview questions were done with the help of two professionals who had done similar interviews before. This ensured that the objectivity of the questions was maintained. Therefore, the thesis serves as a guide for a similar types of companies and situations. The downside of the semi-structured interview is that the subjectivity has to be filtered out from the results. The feedback was analyzed as objectively as possible. The subjectivity of the feedback was reduced by using clear definitions and using well-informed persons in the interviews. There was also an issue of interviewees understanding the questions differently, which was also observed in the interviews. Therefore, it was crucial to make questions understandable by providing guides to the interviewees. Although some downsides appear, the open environment encourages people to speak more honestly about the situation.

The validity, as Saunders et al. (2019, p. 214) refers to the appropriateness of the used measures, accurate analysis of the results, and able to generalize the findings. The validity of the thesis could be increased by executing more performance measurement implementation processes in different organizations. By doing more implementation processes in different fields, observations about the similarities and differences could be obtained. Different needs of the internal stakeholders would certainly incorporate differences. Similarities could be observed definitely. Strategy and objectives are at the key role when designing performance measurement. Thus, similarities between different fields could be observed. The validity of the research was increased by the fact that the performance measurement system was created, and measures were tested in practice.

Best measurement practices and implementation practices were used to improve the validity. Therefore, it can be said that the validity of the study was good.

The performance measurement implementation process was limited to direct procurement. Therefore, more studies are recommended to have a broader understanding of the indirect procurement implementation as well. In addition, quality and supplier evaluation were out of the scope of this study. To widen the understanding of the implementation process, these areas should be researched more.

5.4 Future development

Performance measurement research has a wide spectrum and is in an upward trend due to ever-increasing digitalization. There is still research to be made in the performance measurement sector. Performance measurement literature has still uncovered areas, which require solving. There is a need to bridge a gap between the literature and the practice (Papakiriakopoulos & Pramatari, 2010). In the literature, there is a wide range of frameworks to design performance measurement systems, but fewer about the implementation process (Bourne et al., 1999). Empirical studies are needed to research how performance measurement literature is implemented in different organizations.

For future development, the modified frameworks and proposed solutions should be tested in other empirical performance measurement implementation research. An ideal situation would be to implement performance measurement in other business lines inside the case company. In this case, the organization would stay the same and some strategic goals would stay the same. This would ease the development phase. Another approach would be to implement performance measurement in a similar field.

REFERENCES

Abela, A. (2008). *Advanced Presentations by Design* [Book]. Center for Creative Leadership.

Ahola, T., Laitinen, E., Kujala, J., & Wikström, K. (2008). Purchasing strategies and value creation in industrial turnkey projects [Article]. *International Journal of Project Management*, *26*(1), 87–94. https://doi.org/10.1016/j.ijproman.2007.08.008

Allio, M. K. (2012). Strategic dashboards: designing and deploying them to improve implementation [Article]. *Strategy & Leadership*, *40*(5), 24–31. https://doi.org/10.1108/10878571211257159

Bäckstrand, J., Suurmond, R., Raaij, E., & Chen, C. (2019). Purchasing process models: Inspiration for teaching purchasing and supply management [Article]. *Journal of Purchasing and Supply Management*, *25*(5).

Bello, D. C., & Gilliland, D. I. (1997). The Effect of Output Controls, Process Controls, and Flexibility on Export Channel Performance [Article]. *Journal of Marketing*, *61*(1), 22. https://doi.org/10.2307/1252187

Bourne, M., Mills, J., Neely, A., Platts, K., Wilcox, M., Hamblin, D., & Bicheno, J. (1999). Performance measurement system design: testing a process approach in manufacturing companies [Article]. *International Journal of Business Performance Management*, *1*(2), 154–170. https://doi.org/10.1504/IJBPM.1999.004435

Bourne, M., Mills, J., Wilcox, M., Neely, A., & Platts, K. (2000). Designing, implementing and updating performance measurement systems [Article]. *International Journal of Operations & Production Management*, *20*(7), 754–771. https://doi.org/10.1108/01443570010330739

Brady, T., Davies, A., & Gann, D. M. (2005). Creating value by delivering integrated solutions [Article]. *International Journal of Project Management*, *23*(5), 360–365. https://doi.org/10.1016/j.ijproman.2005.01.001

Brewer, P. C., & Speh, T. W. (2001). Adapting the balanced scorecard to supply chain management [Article]. *Supply Chain Management Review*, *5*(2), 48.

Brignall, S., & Ballantine, J. (1996). Performance measurement in service businesses revisited [Article]. *International Journal of Service Industry Management*, *7*(1), 6–31. https://doi.org/10.1108/09564239610109393

Caniato, F., Luzzini, D., & Ronchi, S. (2014). Purchasing performance management systems: an empirical investigation [Article]. *Production Planning & Control*, *25*(7), 616–635. https://doi.org/10.1080/09537287.2012.743686

Casson, M. (2013). Economic Analysis of International Supply Chains: An Internalization Perspective [Article]. *The Journal of Supply Chain Management*, *49*(2), 8–13. https://doi.org/10.1111/jscm.12009

Chen, C. (2010). Information visualization [Article]. *Wiley Interdisciplinary Reviews. Computational Statistics*, *2*(4), 387–403. https://doi.org/10.1002/wics.89

Chen, I. J., & Paulraj, A. (2004). Understanding supply chain management: critical research and a theoretical framework [Article]. *International Journal of Production Research*, *42*(1), 131–163. https://doi.org/10.1080/00207540310001602865

Cordell, A. (2018). *The category management handbook* (I. Thompson, Ed.) [Book]. Routledge.

David, J. S., Hwang, Y., Pei, B. K. W., & Reneau, J. H. (1999). The impact of purchasing on financial performance. *Arisona State Univesity*.

Dimitri, N. (2013). “Best value for money” in procurement [Article]. *Journal of Public Procurement*, *13*(2), 149–175. https://doi.org/10.1108/jopp-13-02-2013-b001

Dumond, E. J. (1994). Moving Toward Value-Based Purchasing [Article]. *International Journal of Purchasing and Materials Management*, *30*(1), 2–7. https://doi.org/10.1111/j.1745-493X.1994.tb00184.x

Essig, M. (2011). Purchasing and Supply Chain Management: Analysis, Strategy, Planning and Practice, 5th ed., Arjan van Weele. Cengage Learning, Andover, UK (2010). ISBN: 978-1-4080-1896-5 [Article]. *Journal of Purchasing and Supply Management*, *17*(2), 143–144. https://doi.org/10.1016/j.pursup.2011.04.004

Faes, W., Matthyssens, P., & Vandenbempt, K. (2000). The Pursuit of Global Purchasing Synergy [Article]. *Industrial Marketing Management*, *29*(6), 539–553. https://doi.org/10.1016/S0019-8501(00)00127-9

Forslund, H., & Jonsson, P. (2007). Dyadic integration of the performance management process: A delivery service case study [Article]. *International Journal of Physical Distribution & Logistics Management*, *37*(7), 546–567. https://doi.org/10.1108/09600030710776473

Forslund, H., & Jonsson, P. (2009). Obstacles to supply chain integration of the performance management process in buyer-supplier dyads: The buyers’ perspective [Article]. *International Journal of Operations & Production Management*, *29*(1), 77–95. https://doi.org/10.1108/01443570910925370

Forslund, H., & Jonsson, P. (2010). Integrating the performance management process of on-time delivery with suppliers [Article]. *International Journal of Logistics*, *13*(3), 225–241. https://doi.org/10.1080/13675561003712799

Giunipero, L. (2000). Strategic procurement management in the 1990’s: Concepts and cases [Article]. *Journal of the Academy of Marketing Science*, *28*(3), 449–450.

Ha, S. H., & Krishnan, R. (2008). A hybrid approach to supplier selection for the maintenance of a competitive supply chain [Article]. *Expert Systems with Applications*, *34*(2), 1303–1311. https://doi.org/10.1016/j.eswa.2006.12.008

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A Tour through the Visualization Zoo: A survey of powerful visualization techniques, from the obvious to the obscure [Article]. *ACM Queue*, *8*(5), 20–30. https://doi.org/10.1145/1794514.1805128

Heikkilä, J., Kaipia, R., & Ojala, M. (2018). *Purchasing category management: providing integration between purchasing and other business functions* [Article].

Hesping, F. H., & Schiele, H. (2015). Purchasing strategy development: A multi-level review [Article]. *Journal of Purchasing and Supply Management*, *21*(2), 138–150. https://doi.org/10.1016/j.pursup.2014.12.005

Jääskeläinen, A. (2018). Comparison of performance measurement in different purchasing and supply management practices [Article]. *International Journal of Productivity and Performance Management*, *67*(8), 1290–1309. https://doi.org/10.1108/IJPPM-06-2017-0148

Jääskeläinen, A., & Roitto, J.-M. (2016). Visualization techniques supporting performance measurement system development [Article]. *Measuring Business Excellence*, *20*(2), 13–25. https://doi.org/10.1108/MBE-09-2014-0032

Järvenpää, S., & Dickson, G. (1988). Graphics and managerial decision making: research-based guidelines [Article]. *Communications of the ACM*, *31*(6), 764–774. https://doi.org/10.1145/62959.62971

Jiang, Z. (2017). *Procurement management in the supply chain environment: a practical guide to understanding procurement management and enhancing procurement effectiveness and efficiency* [Book]. Chartridge Books Oxford.

Kaplan, R. S. (1996). *The balanced scorecard: translating strategy into action* (D. P. (David P.) Norton, Ed.) [Book]. Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard--measures that drive performance [Article]. *Harvard Business Review*, *70*(1), 71–79.

Keim, D. A. (2002). Information visualization and visual data mining [Article]. *IEEE Transactions on Visualization and Computer Graphics*, *8*(1), 1–8. https://doi.org/10.1109/2945.981847

Kelemen, M. L., & Rumens, N. (2008). *An Introduction to Critical Management Research* [Book]. SAGE Publications.

Kennerley, M., & Neely, A. (2003). Measuring performance in a changing business environment [Article]. *International Journal of Operations & Production Management*, *23*(2), 213–229. https://doi.org/10.1108/01443570310458465

Kerzner, H. R. (2013). *Project Management Metrics, KPIs, and Dashboards* [Book]. John Wiley & Sons, Incorporated.

Ketchen, D. J., & Giunipero, L. C. (2004). The intersection of strategic management and supply chain management [Article]. *Industrial Marketing Management*, *33*(1), 51–56. https://doi.org/10.1016/j.indmarman.2003.08.010

Komatina, N., Nestić, S., & Aleksić, A. (2019). Analysis of the performance measurement models according to the requirements of the procurement business process [Article]. *International Journal of Industrial Engineering and Management*, *10*(3), 211–218. https://doi.org/10.24867/IJIEM-2019-3-241

Kraljic, P. (1983). Purchasing must become supply management [Article]. *Harvard Business Review*, *61*(5), 109.

Larson, E. W., & Gobeli, D. H. (1987). Matrix Management: Contradictions and Insights [Article]. *California Management Review*, *29*(4), 126–138. https://doi.org/10.2307/41162135

Lau, A., & vande Moere, A. (2007). Towards a Model of Information Aesthetics in Information Visualization [Proceeding]. *2007 11th International Conference Information Visualization (IV '07)*, 87–92. https://doi.org/10.1109/IV.2007.114

Lengler, R., & Eppler, M. J. (2007). Towards a periodic table of visualization methods of management [Proceeding]. *Proceedings of the IASTED International Conference on Graphics and Visualization in Engineering, GVE 2007*, 83–88.

Leopold, K. (2015). *Kanban and change leadership: creating a culture of continuous improvement* (S. Kaltenecker, Ed.) [Book]. Wiley.

Lewandowsky, S., & Spence, I. (1989). Discriminating Strata in Scatterplots [Article]. *Journal of the American Statistical Association*, *84*(407), 682–688. https://doi.org/10.1080/01621459.1989.10478821

Lewis, P., Saunders, M. N. K., & Thornhill, A. (2019). *Research methods for business students* [Book]. Pearson Education.

Lohman, C., Fortuin, L., & Wouters, M. (2004). Designing a performance measurement system: A case study [Article]. *European Journal of Operational Research*, *156*(2), 267–286. https://doi.org/10.1016/S0377-2217(02)00918-9

Louviere, J. J., Buckley, P. J., & Devinney, T. M. (2007). Do Managers Behave the Way Theory Suggests? A Choice-Theoretic Examination of Foreign Direct Investment Location Decision-Making [Article]. *Journal of International Business Studies*, *38*(7), 1069–1094. https://doi.org/10.1057/palgrave.jibs.8400311

Luzzini, D., Caniato, F., & Spina, G. (2014). Designing vendor evaluation systems: An empirical analysis [Article]. *Journal of Purchasing and Supply Management*, *20*(2), 113–129. https://doi.org/10.1016/j.pursup.2014.03.002

Malina, M. A., & Selto, F. H. (2004a). Choice and change of measures in performance measurement models [Article]. *Management Accounting Research*, *15*(4), 441–469. https://doi.org/10.1016/j.mar.2004.08.002

Malina, M. A., & Selto, F. H. (2004b). Choice and change of measures in performance measurement models [Article]. *Management Accounting Research*, *15*(4), 441–469. https://doi.org/10.1016/j.mar.2004.08.002

Neely, A. D. (2002). *Business performance measurement: theory and practice* (A. D. (Andy D.) Neely, Ed.) [Book]. Cambridge University Press.

Neely, A., Gregory, M., & Platts, K. (2005). Performance measurement system design: A literature review and research agenda [Article]. *International Journal of Operations & Production Management*, *25*(12), 1228–1263. https://doi.org/10.1108/01443570510633639

Neely, A., Mills, J., Platts, K., Gregory, M., & Richards, H. (1994). Realizing Strategy through Measurement [Article]. *International Journal of Operations & Production Management*, *14*(3), 140–152. https://doi.org/10.1108/01443579410058603

Neely, A., Mills, J., Platts, K., Richards, H., Gregory, M., Bourne, M., & Kennerley, M. (2000). Performance measurement system design: developing and testing a process-based approach [Article]. *International Journal of Operations & Production Management*, *20*(10), 1119–1145. https://doi.org/10.1108/01443570010343708

Neely, A., Richards, H., Mills, J., Platts, K., & Bourne, M. (1997). Designing performance measures: a structured approach [Article]. *International Journal of Operations & Production Management*, *17*(11), 1131–1152. https://doi.org/10.1108/01443579710177888

Nollet, J., Ponce, S., & Campbell, M. (2005). About “strategy” and “strategies” in supply management [Article]. *Journal of Purchasing and Supply Management*, *11*(2), 129–140. https://doi.org/10.1016/j.pursup.2005.10.007

Papakiriakopoulos, D., & Pramatari, K. (2010). Collaborative performance measurement in supply chain [Article]. *Industrial Management + Data Systems*, *110*(9), 1297–1318. https://doi.org/10.1108/02635571011087400

Parmenter, D. (2020). *Key performance indicators: developing, implementing, and using winning KPIs* (Fourth edition.) [Book]. Wiley.

Petersen, K. J., Prayer, D. J., & Scannell, T. v. (2000). An Empirical Investigation of Global Sourcing Strategy Effectiveness [Article]. *The Journal of Supply Chain Management*, *36*(1), 29–38. https://doi.org/10.1111/j.1745-493X.2000.tb00075.x

Pohl, M., & Förstl, K. (2011). Achieving purchasing competence through purchasing performance measurement system design—A multiple-case study analysis [Article]. *Journal of Purchasing and Supply Management*, *17*(4), 231–245. https://doi.org/10.1016/j.pursup.2011.04.001

Pringle, C. D., & Longenecker, J. G. (1982). The Ethics of MBO [Article]. *The Academy of Management Review*, *7*(2), 305–312. https://doi.org/10.2307/257311

Purdy, L., Astad, U., & Safayeni, F. (1994). Perceived Effectiveness of the Automotive Supplier Evaluation Process [Article]. *International Journal of Operations & Production Management*, *14*(6), 91–103. https://doi.org/10.1108/01443579410062112

Purdy, L., & Safayeni, F. (2000). Strategies for supplier evaluation: a framework for potential advantages and limitations [Article]. *IEEE Transactions on Engineering Management*, *47*(4), 435–443. https://doi.org/10.1109/17.895339

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring Organizational Performance: Towards Methodological Best Practice [Article]. *Journal of Management*, *35*(3), 718–804. https://doi.org/10.1177/0149206308330560

Rozemeijer, F. (2000). How to manage corporate purchasing synergy in a decentralised company? Towards design rules for managing and organising purchasing synergy in decentralised companies [Article]. *European Journal of Purchasing & Supply Management*, *6*(1), 5–12. https://doi.org/10.1016/S0969-7012(99)00034-9

Schmidt, G., & Wilhelm, W. E. (2000). Strategic, tactical and operational decisions in multi-national logistics networks: A review and discussion of modelling issues [Article]. *International Journal of Production Research*, *38*(7), 1501–1523. https://doi.org/10.1080/002075400188690

Schuh, C., & Bremicker, M. (2005). *Der Einkauf als Margenmotor: Methoden zur Kostensenkung Mit Fallbeispielen* [Book]. Gabler Verlag.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as Aids to Knowledge Construction [Article]. *Journal of Educational Psychology*, *91*(4), 690–702. https://doi.org/10.1037/0022-0663.91.4.690

Stolle, M. A. (2008). *From Purchasing to Supply Management A Study of the Benefits and Critical Factors of Evolution to Best Practice* (1st ed. 2008.) [Book]. Gabler Verlag. https://doi.org/10.1007/978-3-8349-9695-4

Trautmann, G., Bals, L., & Hartmann, E. (2009). Global sourcing in integrated network structures: The case of hybrid purchasing organizations [Article]. *Journal of International Management*, *15*(2), 194–208. https://doi.org/10.1016/j.intman.2008.09.001

Tully, S. (1995). Purchasings new muscle [Article]. *Fortune*, *131*(3), 75-.

Türkay, M., Saraçoğlu, Ö., & Arslan, M. C. (2016). Sustainability in Supply Chain Management: Aggregate Planning from Sustainability Perspective [Article]. *PloS One*, *11*(1), e0147502–e0147502. https://doi.org/10.1371/journal.pone.0147502

van Looy, A., & Shafagatova, A. (2016). Business process performance measurement: a structured literature review of indicators, measures and metrics [Article]. *SpringerPlus*, *5*(1), 1–24. https://doi.org/10.1186/s40064-016-3498-1

van Raaij, E. (2016). *Purchasing value: purchasing and supply management's contribution to health service performance*.

Vitale, M. R., & Mavrinac, S. C. (1995). How effective is your performance measurement system? [Article]. *Management Accounting (New York, N.Y.)*, *77*(2), 43.

Weele, A. J. van. (2014). *Purchasing & supply chain management: analysis, strategy, planning and practice* (6th edition.) [Book]. Cengage Learning.

Weele, A. J. van. (2018). *Purchasing and supply chain management* (Seventh edition.) [Book]. Cengage.

Woody, K. E. (2012). Conflict minerals legislation: the SEC's new role as diplomatic and humanitarian watchdog [Article]. *Fordham Law Review*, *81*(3), 1315.

Yin, R. K. (2018). *Case study research and applications: design and methods* (R. K. Yin, Ed.; Sixth edition.) [Book]. SAGE.

APPENDIX A: CALCULATION LOGIC OF BUSINESS UNIT DASHBOARD

PAP Procurement dashboard

Spend

The Spend measure is calculated using invoice amounts, not purchase orders.

Default Page level filter selections:

1) Supplier Internal/External: external suppliers
2) View: Periodic
3) Currency Type: Reporting currency (EUR with daily rates)
4) Business Unit: PAP

BU Dashboard

SLICERS

Suppliers:

- Function: User can use the slicer on selecting and searching certain supplier(s)
- Selection:
  - User can select from external/internal suppliers
  - User can select many suppliers or exclude ones

Teams:

- Function: User can select several or individual teams in the organization 3 -level

Legal Company:

- Function: User can select based on legal company's regional location

Business Unit:

- Function: User can select business liner, business units and technology units from the slicer

Year & Month:

- Function: User can select the timeline that is used in the report. Years and months are present

METRICS

SPEND

Spend by VCS:

- Function: Illustrates the spend on different categories
- Vertical bars:
  - Primary Y-Axis: Based on chosen timeline (Year & Month -slicer) of spend/PO Amount by VCS Category Family Description
  - Secondary Y-Axis: CCC percentage
  - Column values: CCC and NON CCC Spend
- Line: CCC percentage.
  - Measure: CCC % Spend. Calculated by choosing CCC suppliers and dividing it with the spend
- Drill down: From VCS Category Family level to VCS Sub-Category level
- Slicers: All the page slicers apply
- Data: Company procurement dataset

Total Spend (donut chart)

- Function: Illustrates the total CCC spend based on the chosen timeline
- Values: CCC and NON CCC Spend
- Drill down: No drill down
- Slicers: All the page slicers apply
- Data: Company procurement dataset

Spend (stacked column chart)

- Function: Illustrates the spend over chosen timeline (Year & Month -slicer)
- Vertical bars:
  - Primary axis: Year-Month hierarchy
  - Column values: include CCC (green) and NON CCC spend (grey)
- Line values: CCC % for the chosen timeline
- Drill down: Drill down from year level to month level, using drill down functionality
- Slicers: All the page slicers apply
- Data: Company procurement dataset

SAVINGS

Spend:

- Function: Illustrates savings over selected timeline.
- Vertical bars: Yearly/Monthly Savings Amounts
- Line Values: Target for selected BU or whole PAP. Visible when YTD is selected. Target data comes from Savings report.
- Drill down: Drill down to month level using drill down functionality
- Slicers: Can be filtered by BU (Business Unit -slicer) or Legal Company -slicer
- Data: Savings report

Savings % by Spend:

- Function: Illustrates the Savings amount in percentages compared to overall spend.
- Vertical bars: Yearly/Monthly Savings % by Spend
- Measure: Savings % by Spend. Calculated by dividing the savings amount by Spend
- Drill down: Drill down to month level using drill down functionality
- Slicers: Can be filtered by BU (Business Unit -slicer) or Legal Company -slicer
- Data: Savings report

OTD %:

- Function: Illustrates the on-time delivery of suppliers
- Line value: OTD by Year & Month
    - Primary axis: Year-Month hierarchy
- Drill down: Drill down to month level using drill down functionality
- Slicers: All the page slicers apply
- Data: Company procurement dataset

Supplier Classification by Spend:

- Function: Illustrates the distribution of suppliers in different categories
- Vertical bars:
    - Primary Y-Axis: Number of suppliers in different categories. Based on chosen timeline (Year & Month -slicer)
    - Secondary Y-Axis: Shows number of suppliers over 12-month period
    - Column values: According the supplier categories: key supplier, main supplier, supplier, n/a
- Line: Number of suppliers over 12-month period
- Drill down: From VCS Category Family level to VCS Sub-Category level
- Slicers: All the page slicers apply
- Data: Company procurement dataset

Top 15 Suppliers by Spend:

- Function: Illustrates the top 15 supplier by company name and also by corporation name
- Horizontal bars:
    - Primary X-Axis: Spend of suppliers in different categories. Based on chosen timeline (Year & Month -slicer)

    - Column values: According the supplier categories: key supplier, main supplier, supplier, n/a
- Drill down: From corporation name -level to the company name -level
- Slicers: All the page slicers apply
- Data: Company procurement dataset

Buyer and Supplier Table:
- Function: Illustrates values by different organizations in table form. Can be drilled down to supplier and buyer -levels
- Columns:
    - PO (Purchase Order) Receipt Amount, PO Amount, # of PO Lines (Number of PO lines), OTD, OTD Days Early, OTD Days Late, Lead Time, CCC %, Spend
- Drill down: Can be drill down from organization level to the organization level 3. From organization level 3, data can be drill down to purchaser level, from purchaser level can be drill into supplier level, and from supplier level to the individual purchase orders.
- Slicers: All the page slicers apply
- Data: Company procurement dataset

Buttons

Business unit buttons:
- Function: End-user can select different business units and the slicers will filter accordingly
- Functionality:
    - CN:
        - Legal Company -slicer is chosen to be China
        - Business Unit -slicer is PAP
    - PM, SPR, TM
        - Business Unit -slicer is the selected business unit
    - PAP
        - Business Unit -slicer is selected as PAP

Periodic button
- Function: Shows the months and years selected in the Year & Month -slicer
- Functionality: No calculations made

YTD button
- Function: Shows YTD values according to selected timeline in Year & Month -slicer
- Functionality: Sums all the previous months/years into each other until the final date is reached

R12M button
- Function: Filters the Year & Month -slicer to show only the full 12-months from this day
- Functionality: No calculations made