LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT
School of Business and Management
Master's Programme in Supply Management

Julia Laitinen

IMPACT OF SERVICE-LEVEL AGREEMENTS ON DELIVERY RELIABILITY

Examiners: Professor Veli Matti Virolainen
Professor Katrina Lintukangas

ABSTRACT

Author:	Julia Laitinen
Title:	Impact of service-level agreements on delivery reliability
Faculty:	LUT University School of Business and Management
Degree programme:	Master's Programme in Supply Management
Year:	2021
Master's Thesis:	Lappeenranta-Lahti University of Technology, 68 pages, 17 figures, 16 tables
Examiners:	Professor Veli Matti Virolainen Professor Katrina Lintukangas
Keywords:	service-level agreement, delivery reliability, order fulfillment, supplier performance

The main objective in this thesis is to investigate if service-level agreements have an impact on delivery reliability. This thesis is written from the perspective of supplier, a case company, that operates as manufacturer and supplier in the automotive industry. Theoretical framework consists of order fulfillment processes, supplier performance evaluation and risks associated with delivery reliability. Empirical part of this thesis aims at providing insights of where possible fall below the delivery reliability expectations is deriving from and evaluates how case company is performing as supplier in terms of delivery reliability. Empirical research is carried out as a single case study and the research is conducted with quantitative methods. Secondary data for statistical analyses is gathered from case company internal sources. From the dataset two sample groups are formed based on if there is a service level agreement with supplier (case company). Statistical analyses compare the actual delivery reliability figures of these two groups to each other. Results show that there is statistical difference between the two groups, and the impact of having a service-level agreement is not in the positive direction. This case study showed that customers without service level agreement did have a higher delivery reliability over the review horizon of 2019-2020. The reasons behind the findings are analyzed further and corrective actions are suggested for the case company.

TIIVISTELMÄ

Tekijä:	Julia Laitinen
Tutkielman nimi:	Palvelutasosopimusten vaikutus toimitusvarmuuteen
Tiedekunta:	Kauppatieteellinen tiedekunta
Koulutusohjelma:	Hankintojen johtaminen
Vuosi:	2021
Pro Gradu-tutkielma:	LUT-Yliopisto
	68 sivua, 17 kaaviota, 16 taulukkoa
Tarkastajat:	Professori Veli Matti Virolainen
	Professori Katrina Lintukangas
Avainsanat:	palvelutasosopimus, toimitusvarmuus, tilausten toteutuminen, toimittajan suorituskyky

Tämän tutkielman tavoitteena on selvittää, onko palvelutasosopimuksilla vaikutusta toimitusvarmuuteen. Tutkimuksen kohteena on case-yritys, joka toimii valmistajana ja toimittajana autoteollisuuden toimialalla. Tämä työ on kirjoitettu toimittajayrityksen näkökulmasta. Teoreettinen viitekehys koostuu tilausten täyttämisprosesseista, toimittajien suorituskyvyn arvioinnista ja toimitusvarmuuteen liittyvistä riskeistä. Empiirisen tutkimuksen tavoitteena on selvittää, mistä toimitusvarmuusodotusten alittuminen johtuu sekä arvioida case-yritystä toimittajana toimitusvarmuuden näkökulmasta. Tutkimus toteutettiin tapaustutkimuksena, kvantitatiivista menetelmää hyödyntäen. Tiedot tilastollisia analyyseja varten on kerätty case-yrityksen sekundaarisista lähteistä. Aineistosta muodostettiin kaksi otantaryhmää sen perusteella, onko toimittajan (case-yrityksen) kanssa olemassa palvelutasosopimus vai ei. Tilastollisissa analyyseissä verrattiin näiden kahden ryhmän todellisia toimitusvarmuuslukuja toisiinsa. Tulokset osoittavat, että näiden kahden ryhmän välillä on tilastollisia eroja toimitusvarmuuden suhteen. Tämä tapaustutkimus osoitti, että otantaryhmällä, joilla ei ollut palvelutasosopimuksia käytössä case-yrityksen kanssa, toimitusvarmuus oli korkeampi tarkastelujaksolla 2019--2020. Havaintojen syitä analysoidaan tarkemmin ja näiden pohjalta case-yritykselle ehdotetaan korjaavia toimenpiteitä.

ACKNOWLEDGEMENTS

Two years studying at LUT have been valuable for both personal and professional growth. I consider applying, and luckily getting admitted, to LUT one of the best decisions I have made. Time went by quickly and although 2020 was the year of home studying, I am very glad that we got have the normal university life experience before that.

I wish to express my gratitude to my supervisors Professor Veli Matti Virolainen and Professor Katrina Lintukangas for your support along the master's thesis process. I consider your valuable advices and encouraging words especially at the beginning of the process being a great success factor for completing the work. For me master's thesis was an interesting task and a true learning experience.

A big thanks goes to my family and closest friends for your continuous support. You helped me stay focused, encouraged me when I needed support, sat with me on Saturday's studying and reminded me that as soon as I complete this thesis, the sooner I get to enjoy upcoming summer.

In Helsinki 25.5.2021

Julia Laitinen

TABLE OF CONTENTS

1 INTRODUCTION 1
1.1 Background of the study 1
1.2 Research objectives and research questions 2
1.3 Research methodology 4
1.4 Limitations 5
1.5 Theoretical framework 7
1.6 Key concepts 8
1.7 Outline of the thesis 9
2 LITERATURE REVIEW 11
2.1 Service-level agreements 11
2.1.1 Order fulfillment as integral supplier operations process 15
2.1.2 Manufacturing models 20
2.2 Assessing delivery reliability 21
2.2.1 Measuring supplier performance 23
2.2.2 Supplier evaluation and development 27
2.2.3 Risks affecting delivery reliability 30
3 METHODOLOGY 33
3.1 Research design 33
3.2 Data collection 35
3.2.1 Sampling 38
3.3 Analyses design and hypothesizing 40
3.3.1 T-test 41
3.3.2 Spearman’s correlation 42
3.4 Research reliability and validity 43
4 EMPIRICAL FINDINGS 46
4.1 Descriptive tables 46
4.2 Testing for equal means across groups 50
4.2.1 Delivery reliability across groups 50
4.2.2 Supplier delivery reliability across groups 52
4.2.3 Warehouse delivery reliability across groups 53
4.3 Testing for correlation between variables 55
4.4 Summary of the results 57
5 DISCUSSION 59
5.1 Answers to research questions 59

5.2 Conclusions 66
5.3 Managerial recommendations 67
5.4 Limitations and suggestion for future research 67
LIST OF REFERENCES 69

LIST OF FIGURES

Figure 1. Scope of this thesis in the context of internal supply chain
Figure 2. Theoretical framework
Figure 3. Operational level order fulfillment processes
Figure 4. Order fulfillment process
Figure 5. SCOR model of delivery reliability
Figure 6. Supplier performance measurement system
Figure 7. Life cycle of performance measurement system
Figure 8. Kraljic's purchasing portfolio matrix
Figure 9. Reseach design steps
Figure 10. Data collection stages
Figure 11. Conceptual framework
Figure 12. Strength of correlation
Figure 13. Missed on-time deliveries during observation months
Figure 14. Delivery reliability figures, for all and based on grouping
Figure 15. Non-normal distribution of variable order lines total
Figure 16. Non-normal distribution of variable missed deliveries total
Figure 17. Conceptual framework revised with results

LIST OF TABLES

Table 1. Outline of this thesis
Table 2. Examples of industries operating under service-level agreement constrains
Table 3. Strategical level order fulfillment processes
Table 4. Long panel data matrix
Table 5. Customers in the raw data prior to sampling
Table 6. Sample groups
Table 7. Descriptive table for all observations
Table 8. Descriptive table by groups
Table 9. Part one: Test of equality of variances for variable delivery reliability
Table 10. Part two: T-test of equal means for variable delivery reliability across groups
Table 11. Part one: Test of equality of variances for supplier delivery reliability

Table 12. Part two: T-test of equal means of variable supplier delivery reliability across groups
Table 13. Part one: Test of equality of variance for warehouse delivery reliability
Table 14. Part two: T-test of equal means of variable warehouse delivery reliability across groups
Table 15. Shapiro-Wilk test for normal distribution
Table 16. Spearman’s correlation

1 INTRODUCTION

1.1 Background of the study

Performance of the supplier has received significantly higher emphasis in the supplier-buyer relationship management in recent years (Maestrini et al., 2018, 2040). Being able to quicky respond to customer requirements and serving the customer in the uncertain supply chain together with fluctuating demand, are competitive factors in today's business (Christou & Ponis, 2009, 3063). According to Park et al. (2010, 505) supplier performance is recognized as a central parameter in supplier evaluation actions, together with supplier capability and collaboration.

At the center of this study are the concepts of service-level agreement (SLA) and supplier performance in terms of delivery reliability and the relationship of these two to each other. This thesis is about detecting possible linkages between service-level agreements use and delivery reliability performance. To find out what is the difference of delivery reliability for customer accounts where service-level agreement is determining operations compared to those where such agreements are not in use. The focus is on the supplying company point of view with the aim to give alternative perspective to the topics around supplier performance.

This thesis is written for a company operating in the automotive industry, here referred to as the case company. The business division of the case company where this research is conducted delivers tens of thousands of order lines in a year to customers Europe wide. Key performance indicator of customer delivery reliability determines everyday operations. Customer expectations in terms of delivery reliability level are in focus on the management level. Performance is followed systematically both internally and externally on the customer's side and case company is giving great emphasis to it.

Automotive industry has been studied widely in the supply chain management context and is presented as example in many of the academic research around the theme. Much research in the supply management field is interested in this industry due to its complex supply chain networks and multi-level tier supplier operations. Literature around concepts and processes related to service-level agreements is relevantly recently studied in the academic literature. Supplier performance on the other hand is widely investigated topic where a lot of literature exists. Although the research is conducted in the supplying company, this thesis hopes to contribute to the research of combining the two concepts of service-level agreement and supplier performance together and presenting findings that are in the interest of both – the supplying and purchasing company.

Case company will gain valuable information of the importance and impact of engaging into service-level agreements with customers. The aim is to provide an overview of the nature of service-level agreements, how those impact operational processes and to study the connection between service-level agreements and delivery reliability. This thesis aims to provide statistically analyzed data that case company management and whoever's interest it is internally, can utilize for multiple purposes. The research can also be viewed from the self-evaluation aspect. As the intention is to investigate, compare and evaluate delivery reliability of the company, case company will as well get a comprehensive understanding of how well they are performing as supplier to their customers. It is expected that case company management would benefit from the comprehensive academic literature around the topic and as well of the results of the empirical research in decision making i.e., for internal processes evaluation purposes and for when considering motives and benefits of service-level agreements.

1.2 Research objectives and research questions

To understand the scope of this thesis, its objectives, topics, and perspectives of the research better, Figure 1 is created and explained open in this chapter. Traditional way to consider internal supply chain of an organization considers functions for purchasing, production, and distribution. In this thesis context the distribution function is instead specified to consider order fulfillment management. The thesis considers two flows: operational flow between order fulfillment operations and customer, but as well the success of internal supply chain processes. The relationship focuses are highlighted in the Figure 1.

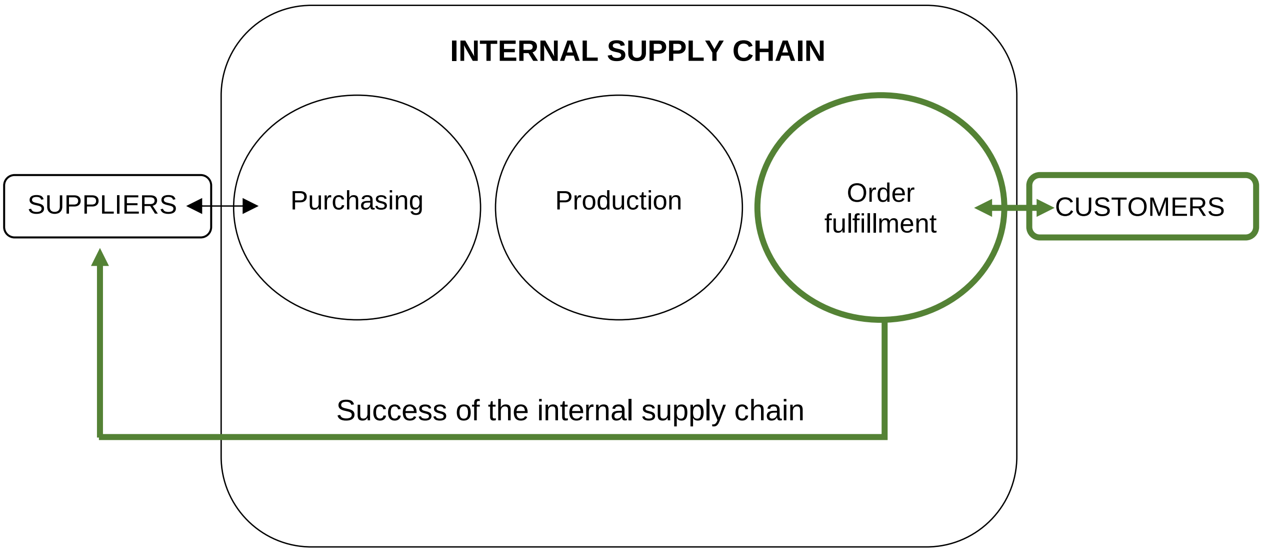


Figure 1. Scope of this thesis in the context of internal supply chain.

The concept of service-level agreement is evaluated from supplier perspective and processes covered are related to order fulfillment. Delivery reliability on the other hand, is considered to have dual meaning, when talking about the perspective. Primarily, delivery reliability is important for the customer of the case company, who determines sufficient delivery reliability level. Secondly, it is important for the case company, whose delivery reliability towards its customer, is reflected by the good delivery reliability of its own suppliers. To scatter the meaning of delivery reliability into two pieces, it is stated, that delivery reliability of the case company towards its customers reflects two things, (1) how well order fulfillment management is conducted and (2) how well the suppliers of the case company are succeeding in delivering the goods against the orders on time. Therefore, in this thesis context, delivery reliability can be considered to affect and be in the interest both, the customer of the case company and the case company as customer to its suppliers.

The main objective in this thesis, is to find out if service-level agreements have an impact on the delivery reliability. To find out how the case company is performing as supplier and to identify possible connection between service-level agreement and delivery reliability. This objective is approached by setting one main research question (RQ) supported by two sub-research questions (S-RQ). As well with the support of theoretical background on the literature review, the aim is to via empirical research, answer the research questions at the last discussion chapter of this thesis. Research question and sub-research questions are formulated as follows:

Main research question (RQ)
What is the impact of service-level agreement to delivery reliability?

Sub-research question 1
What is the cause of falling below delivery reliability expectations?

Sub-research question 2
How is the case company performing as supplier in terms of delivery reliability?

There are relevantly recent studies regarding service-level agreements in the supply chain management research. Agreed by Kloos & Pibernik (2020, 204) who argue that service-level contracts and service-level agreements have received significantly increasing interest in the research of operations management. As well it seems that service-level agreement is rarely selected as master's thesis topic or yet, as main concept. Therefore, it feels important to conduct this study with delivery reliability and service-level agreements as the main concepts and to look the impact of both, the supplying company and purchasing company.

1.3 Research methodology

Empirical research is conducted as quantitative case study in one organization setting. A single-case study approach is selected for this thesis due its suitability and with the aim to have deeper insights of the case company operations. Case study method is especially suitable for operations management study context (Voss et al., 2002, 195). Research is carried out by quantitative methods and the quantitative data is analyzed with statistical analysis. Quantitative data of this research is numerical data, that presents five different measured variables. (Saunders et al., 2016, 500).

The study utilizes archival and documentary research approach, where the secondary data is gathered from company internal databases, websites, and archives. Data collection is performed by analyzing customer contracts and agreements from company archives and by generating reports from case company internal databases. Secondary sources utilized are delivery reliability reports, contracts, agreements, and other relevant documents. Based on the data collection two sample groups are formed for later analysis purposes. Group 1 where customers have service-level agreement in place and Group 2 where such agreements are not in use.

The objective of this research is to investigate possible linkages between service-level agreements and delivery reliability. The best way to do this by statistical analysis and hypothesis setting. A total of four hypotheses are tested in the empirical part of this thesis. From the data set, a descriptive table for the five measures is created, and discussed. For the comparison between the groups (customers who have SLA and those who do not), a t-test is selected as suitable for the mean comparison purpose. T-test compares means of two groups and shows if the difference between these groups is statistically significant (Geher & Hall, 2014, 235, 248). For the fourth hypothesis a correlation analysis between two variables is conducted. Statistical analyses selection is justified in depth in Chapter 4 and results are presented in Chapter 5. Final discussion chapter aims at responding to the research questions in the light of comprehensive literature review and empirical research results.

1.4 Limitations

Service-level agreements consist of different components, that relate to for example quality requirements, price, and order quantity to name a few. The content of the agreements varies by customer or is determined by company policies. First limitation is that this research focuses solely on the service-level agreement component of delivery reliability. Intention is to find out if service-level agreements impact performance of the supplier specifically in terms of delivery reliability expectations. For example, quality related measures are not covered in this thesis context.

Second justification refers to the perspective to implement the research from supplying company point of view. In general, it can be said that the products case company’s customers source from the case company are of the same product portfolio. Therefore, it is seen as logical to do a research where sourced goods are at least to some extent, same. When compared to the purchasing company point of view, if the research would be implemented from that perspective, sourced goods would represent different product categories sourced from different suppliers.

Third limitation concerns the topics chosen for the literature review of this thesis. Lambert & Schwieterman (2012) present eight macro supply chain management processes. This thesis bases to their proposal of the management processes and focuses on two of them: order fulfillment and supplier relationship management. The operations management topics chosen and covered in this thesis are order fulfillment, written from the perspective of supplier and supplier evaluation, from the perspective of buyer. For example, the process of demand management is not covered in this thesis. Although it is relevant, it is considered far strategical and focusing on the decisions before the customer orders are received i.e., in the capability of supply chain, marketing requirements and ways to synchronize supply and demand the best way. (Lambert & Schwieterman, 2012, 351). Naturally, as demand management affects the order fulfillment and has mutual interfaces there are some processes that can apply to both SCM sub-processes.

When discussing the manufacturing models, it is worth mentioning that make-to-stock model is not covered that extensively, due to the relevance of SLA and allocation planning to specifically interesting for make-to-order manufacturing models. As well, case company as supplier operates mostly MTO basis. What comes to supplier relationship management (SRM), the focus is on supplier evaluation and performance measurement.

To consider empirical research limitations, there are some aspects left out of the scope. This study as case study – to be more specific, one division of case company – focuses only in one organization setting. Therefore, the customer's selected for the empirical analysis, are very narrow sampling of case company's complete customer base. Customers selected for the analysis are (1) operating in Europe (2) are spare-part or manufacturing customers and (3) have been actively ordering during the research review horizon of 2019-2020. Empirical research is restricted to the use of quantitative research methods, and therefore open questions are not covered in this research but are rather presented as ideas for further research. This also means that empirical research does not seize into detailed root causing on order line or customer level, in cases of analyzing possible poor delivery reliability.

1.5 Theoretical framework

Theoretical framework for this thesis is formed based on literature review around the topics. Main theories are selected based on Lambert (2008) suggestion of the eight macro-processes of supply chain management, where order fulfillment and supplier relationship management are two of the eight processes. Latter one is specified for this theoretical framework to specifically consider supplier performance evaluation.

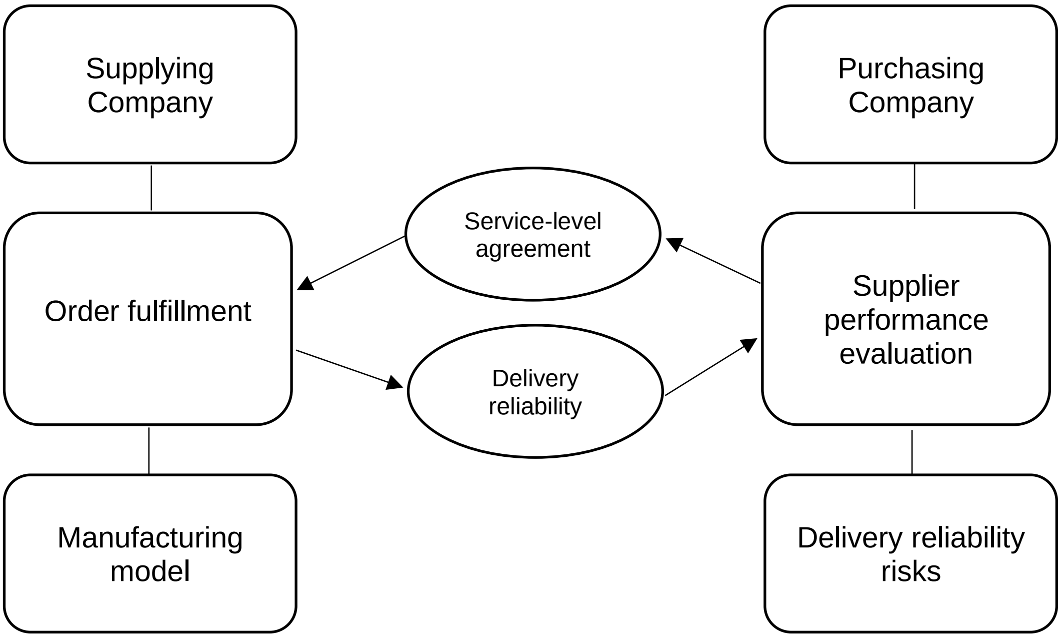


Figure 2. Theoretical framework

On the left-hand side of Figure 2 the perspective of supplying company is presented and the main concept for this is order fulfillment. The selection of manufacturing model affects order fulfillment processes widely. On the right-hand side the purchasing company point of view is presented in the light of the main concept of supplier performance evaluation. Together with delivery reliability affecting risks these two concepts form the base for purchasing company perspective. In between the parties, there are the main concepts under investigation of this thesis: service-level agreement and delivery reliability.

1.6 Key concepts

In this chapter the key concepts are defined and described to ease the reading experience for the reader. Concepts are looked from the theory perspective, particularly relevant to this thesis context. Also, from the case company point of view if information is available. More in-depth analyzes and related concepts are given in next chapters. Key concepts are presented in an order of appearance of the next literature review chapter.

Service-level

Service-level is the percent of deliveries from supplier, delivered on time (Altendorfer & Jodlbauer, 2011, 1827). Synonyms to service-level are item fill rate (Thomas, 2005, 74) and customer service level (CSL) as defined in the case organization. As well delivery reliability indicates same measure than service-level.

Service-level agreement

Service-level agreement is an agreement between the buying company and supplying company, that defines the performance expectations of the supplier (Chen & Thomas, 553, 2018). In the case company these types of contracts and agreements with customer's are referred to as logistics agreements or supply chain alignments.

Service-level agreement customer

A customer or a customer group whose operation are determined by above defined service-level agreement. For the case company service-level agreement customer, later referred to as SLA customer has a high focus from the operational employee and management side. As one of the SLA characteristics, the delivery reliability expectation is followed and monitored on a defined review horizon level from both case company and buying company side. Object for the case company as supplier is to reach the delivery expectation level for each review horizon to the best extent.

Delivery reliability

Delivery reliability in the context of this thesis means an on-time delivery leaving the case company's warehouse to customer, excluding any delay that might occur during transit from supplier warehouse to customer facilities. Similar terms are service-level and delivery accuracy. For this thesis, the term 'delivery reliability' is chosen to describe best above definition, which is as well used in the case company.

Supplier relationship management (SRM)

Supplier relationship management is a management process that covers all actions between the organization and its suppliers (Tseng, 2014, 40). Supplier relationship management consists of processes such as supplier assessment, supplier evaluation and supplier development (Park et al., 2010, 495).

Supplier evaluation

Supplier evaluation as an integral process of supplier relationship management. It reflects and defines how suppliers are selected (Baskaran et al., 2012, 648). As well, supplier evaluation purpose is to provide assessment of the supplier performance.

1.7 Outline of the thesis

This thesis follows the structure presented in Table 1. Introduction chapter outlines background of the study, research objectives and limitations, research methodology and key concepts. After that, main concepts are evaluated and presented in the light of academic literature. Following a chapter of research methodology that includes justification of the research design, data description and collection, sampling, analysis design and evaluation of the research reliability and validity. Fourth chapter presents the empirical findings of the research results. This thesis ends with discussion chapter where the research questions are answered, and further research ideas presented.

Table 1. Outline of this thesis.

Chapter	Contents
CHAPTER 1 Introduction	•Background of the study •Objectives and research questions •Methodology •Limitations •Key concepts
CHAPTER 2 Literature review	•Service-level agreements •Order fulfillment •Manufacturing models • Delivery reliability • Supplier evaluation • Related risks
CHAPTER 3 Methodolody and data collection	•Research design •Data collection and sampling •Analyses selection and hypotheses •Reliability and validity
CHAPTER 4 Empirical findings	•Descriptive tables •Test of equal means across group •Test of correlation across variables •Summary of the results
CHAPTER 5 Discussion	•Answers to research questions •Conclusions •Managerial recommendations •Further research ideas

2 LITERATURE REVIEW

This chapter presents a review of the existing academic literature relevant to this thesis research. Sole focus is on the main concepts of this thesis: service-level agreement and delivery reliability. This chapter is divided into two sections based on the perspective in question. First part, Chapter 2.1 and its sub-chapters presents the perspective of the supplier company, written from the order fulfillment process management point of view. The concepts covered in this chapter are service-level agreements and service-level contracts, order fulfillment, and manufacturing systems. Second part, Chapter 2.2 and its sub-chapters evaluate purchasing function perspective in terms of supplier relationship management. There the focus is on topics of delivery reliability, supplier evaluation and supplier performance measurement and supply chain risks affecting delivery reliability.

2.1 Service-level agreements

Service-level or item fill rate measures the level of customer demand being fulfilled from supplier's stock (Thomas, 2005, 74). Definition by Altendorfer & Jodlbauer (2011, 1827) proposes that service level equals to the percent of orders delivered on time. Alternatively, service-level can be defined as probability under restrictive rules, where the demand is aimed to be fulfilled with the best possible likelihood (Jiang et al., 2019, 2365). When considering the relevance from the inventory management point of view, service-level equals to the fill rate that is the "*fraction of orders filled from stock*" (Hopp et al., 1997, 327-328). This definition is especially used when orders are fulfilled from the existing stock and delays of goods arrival to stock are not considered.

Service-level contracts and service-level agreements have received increasing interest in the research of operations management (Kloos & Pibernik, 2020, 204). Service-level agreement is an agreement between the buying company and supplying company, that defines the performance expectations of the supplier (Chen & Thomas, 553, 2018). Katok et al., (2008, 609) emphasize that service-level agreement is the commitment the supplier takes when entering business with the buying company. According to Kloos & Pibernik (2020, 204) not all authors of the existing literature use the term service-level agreement or service-level contract. As well as in their study, in this literature review and thesis context service-level agreement is defined as an agreement, where certain service-level is expected but the consequences of meeting the expectation are vague or nonexistent. Service-level contract on the other hand, is a binding and written contract where failure to meet the expected level, derives into financial consequences, i.e., penalties. (Kloos & Pibernik, 2020, 204). Focus on this thesis will mostly be on the first, agreement-based service-level expectations.

Both buyer and supplier have their own motives for initiating service-level agreements to determine material or service flow. For buyer, the primary motive relates to enhancing supplier performance, as supplier is motivated to improve their service further than without SLA expectations (Chen & Thomas, 2018, 553). By setting up SLAs with suppliers, buying company is looking for supplier to commit to product availability, good quality of the delivered goods and timely deliveries. In case of poor performance against the SLA, buyer eventually expects contract prices to be redetermined to their benefit. (Liang & Atkins, 2013, 1103). Supplier is motivated to perform on a sufficient level due to possible penalties of SLA or simply to keep the business with the customer (Chen & Thomas, 2018, 553). Penalties from the supplier perspective are excessive and often result as revenue for the buyer companies (Chen, 2018, 1326).

Based on the existing academic literature, companies of many industries have service-level agreements (SLA) in use with their suppliers. Table 2 presents a proportion of industries that have SLAs in use. Intention is to provide an overview of versatile industries using SLAs to strengthen their operatives. Examples of authors and their academic work is mentioned. As well the penalty type or consequence of falling below SLA expectations that is discussed in each article is mentioned.

Table 2. Examples of industries operating under service-level agreement constrains.

INDUSTRY	AUTHOR	CONSEQUENCE
Information technology	Milner, J. M. & Olsen, T. L. (2008) Service-Level Agreements in Call Centers: Perils and Prescriptions. Management science. [Online] 54 (2), 238–252.	Percentile delay penalty Convex penalty
Manufacturing, Production	Sieke, M. A. et al. (2012) Designing Service Level Contracts for Supply Chain Coordination. Production and operations management. [Online] 21 (4), 698–714.	Flat penalty Unit penalty
	Ching, W.-K. et al. (2011) Inducing high service capacities in outsourcing via penalty and competition. International journal of production research. [Online] 49 (17), 5169–5182.	Fixed penalty
Retail	Chen, C. & Thomas, D. J. (2018) Inventory Allocation in the Presence of Service-Level Agreements. Production and operations management. [Online] 27 (3), 553–577.	Flat penalty Unit penalty
	Chen, C.-M. (2018) A review and analysis of service level agreements and chargebacks in the retail industry. The international journal of logistics management. [Online] 29 (4), 1325–1345.	Variable-fee penalty Flat-fee penalty
Agri-food	Baghalian, A. et al. (2013) Robust supply chain network design with service level against disruptions and demand uncertainties: A real-life case. European journal of operational research. [Online] 227 (1), 199–215.	Service-level as competitive factor

In the information technology sector, in call center environments, service-level is defined either between the service provider and client or internally by the service provider, penalties applied can be either delay percent based or convex penalties (Milner & Olsen, 2008, 239). Sieke et al., (2012, 698) article discusses and analyzes service-level contracts in the manufacturing setting. In their study the observed manufacturers operated based on flat penalty or unit penalty contracts. Ching et al. (2011) contributes to the SLA research for make-to-order companies and studies the use of fixed-penalty and the impact of that to supplier incentives. Chen & Thomas (2018, 554) highlight that it is very common to have SLAs with penalty clauses, for most of the SLA's in their study (39 out of 45) penalties were applied. Chen (2018) provides exploratory findings in his study, presenting (1) comprehensive overview of SLAs (2) most typical SLA parameters and (3) SLA design in retail industry context. Baghalian et al. (2013, 207) study a rea-life case from the agri-food (rice) industry. In their study, penalty-based contracts are not at center, but service-level's importance is highlighted and seen as competitive advantage. This last study from agri-food sector is selected as an example to the table to point out the usefulness of service-level based operations in many industry contexts.

Service-level contracts and agreements typically consist of three components:

(1) performance level expected to be achieved
(2) period when performance is reviewed and evaluated (review horizon)
(3) possible financial or other consequences of not meeting the expected performance level over the agreed review horizon (Kloos & Pibernik, 2020, 203).

The performance level is steered by the customer or company internally (Gurvich et al., 2008, 280). Thomas (2005, 79) in his study present two possible time scopes to evaluate the supplier target fill-rate: short review horizon and long review horizon. Chen & Thomas (2018, 553) mention that time horizon can also be measured per each customer order fulfilled. The benefit of short review horizon is that during this time there might not be any large orders to be fulfilled, and this may increase the probability of target fill rate being met. Controversially long review horizons allow the suppliers to manage high peaks and recover from high customer demand. (Thomas, 2005, 79).

Penalty, also referred to as fee, fine, failure cost, violation of agreement or recovery charge is the financial charge supplier is expected to pay to the customer in case of malperformance, agreed in the service-level contract or agreement. Penalty types vary per what is agreed in the contract or agreement. Most typically penalties are in form of variable-fee or flat-fee charges. Variable-fees are determined according to the quantity or percent of nonperformance by the supplier. Then again flat-fee penalties do not influence on how much the target performance is falling below. Flat-fee penalty is fixed in terms of amount of chargeback. (Chen, 2018, 1329, 1337). Convex penalty, as introduced in Milner & Olsen (2008) article in the information technology sector, is a charge that increases nonlinearly and in addition it may contain flat-fee penalty.

Having service-level agreement in place, buying companies expect their suppliers to follow agreed performance levels and other requirements and in case of falling back from these levels in form of late or partial deliveries, supplier to pay back financial charges (Chen, 2018, 1338). Target service-level contributes as incentives for supplier (Liang & Atkins, 2013, 1117). Katok et al., (2008, 623) study confirm that service-level agreement use is effective for supply chain coordination purposes. Ching et al., (2011, 5176-5177) study demonstrates that having fixed-penalty, non-reliant on the level of falling back from the agreed performance level, in fact encourages the suppliers to operate with the lowest capacity possible, due to the penalty fee being relatively low. Therefore, they argue that increase of penalty level in form of using other than fixed penalties is always beneficial for the buying company. Same is noted in Liang & Atkins (2013, 1117) study, where they emphasize that fixed penalty costs diminish supplier efforts once the performance has already fell back from the expected level. According to them, this should be considered when designing service-level agreements. When considering variable-fee performance level deviation, the allowance of fallback from the expected target should be limited. (Liang & Atkins, 2013, 1115-1117).

2.1.1 Order fulfillment as integral supplier operations process

To assess the operations important from the supplier point of view, the management practices need to be put into context of supply chain management. Supply chain management (SCM) is considered as an essential success factor of companies' competitiveness. Since 1980 there has been academic interest and discussion around the topic. (Xiang Li 2014, 1). On the same path is Otto & Kotzab (2003, 306) by emphasizing that SCM is a hot topic in the field of management research. There have been many interpretations when defining the term "supply chain management" and the definition has evolved over time. According to Min et al. (2019, 49) what is still certain today, is that supply chain management is:

- strategical competitive advantage
- a way to create value to customer
- a way to collaborate and coordinate interorganizational

Lambert (2008) introduced the eight macro-processes of supply chain management. The research continued in the Lambert & Schwieterman (2012, 337) article where the processes are comprehensively described. These SCM processes are customer relationship management, supplier relationship management, customer service management, demand management, order fulfillment, manufacturing flow management, product development and commercialization and returns management. Relevant to this thesis, focus on the supplier side operations management is presented via order fulfillment processes. This and related processes are elaborated in-depth in the following chapters. In the later chapters when the perspective of the buying company is presented, the focus is on the supplier relationship management practices.

Order fulfillment is an integral process of the supplying company operations. Okongwu et al. (2012, 581) in their article '*A decision support system for optimising the order fulfilment process'* emphasize the importance of order fulfillment process to an organization. They consider order fulfillment process to be the key factor of customer satisfaction in the competitive business environment. Order fulfillment process is designed customer-centrally but having the internal and supply chain capabilities of the supplying company in mind (Croxton, 2003, 22). According to findings of Brabazon & MacCarthy (2017, 144) manufacturers in for example automotive industry use very similar order fulfillment process in terms of structure and main content. In Kritchanchai & MacCarthy (1999, 818) article, the order fulfillment processes research was divided into two stages: pre-order stage and order receiving stage and sub-processes of those. The division is presented in Figure 3.

Stage 1: pre-order

- forecasting
- capacity planning
- material planning
- stock level planning

Stage 2: order receiving

- order processing
- allocation planning
- order confirmation
- production planning
- inventory management
- product delivery

Figure 3. Operational level order fulfillment processes (Kritchanchai & MacCarthy, 1999, 818).

Stage one being the pre-order stage where all activities possible are finalized before the actual order arrives. These are processes such as forecasting and planning of capacity, materials, and stock levels. Stage two consisting of the activities related to order processing. Besides the actual order processing, this stage consist of order confirmation, allocation planning and setting the priority, production planning, inventory management and finally product delivery. (Kritchanchai & MacCarthy, 1999, 818). In next chapters the concept of allocation planning is further explained.

While the Kritchanchai & MacCarthy (1999) study focused on the operational processes of order fulfilment and the division of those to pre- and post-order receiving stages, Croxton (2003) examined the order fulfilment process having two-levels in terms of strategical and operational processes. These include much more than just delivering the orders. When considering the strategical level, the main objective is to assess and develop the processes to improve the financial performance of the supplying company, its customers', and supply chain partners. Operational level on the other hand considers all the transactions related to order fulfillment, an example presented in the previous Figure 3. Croxton (2003, 19). Table 3 presents the five strategical level process steps where the objective is to create a sustainable order fulfillment strategy.

Table 3. Strategical level order fulfillment processes. Adapted from Croxton (2003, 22).

Review of strategies, structure and customer requirements	•Review internal marketing and supply chain stratgeies •Review customer requirements •Determine supply chain capabilities
Define requirements	•Evaluate O2C cycle •Define customer specific lead time and customer service requirements •Assess core competences
Evaluate supply chain network	•Assess the financial capabilities of current SC setting •Warehouse, supplier and plant locations •Determine transportation modes
Define operational order fulfillment	•Define allocation and prioritization rules •Determine which systems are used in the operational transactions
Develop measurement system	•Determine KPI's •Set target levels

In a perfect situation all customer orders can be fulfilled on time or from the available supply. It is as well strategical level alignment of what to do in a situation where the customer orders cannot be fulfilled and how to solve these situations, keeping the customer-central mindset. (Croxton, 2003, 24). In a situation where not all customer orders can be fulfilled from the available supply, allocation planning comes in question. Allocation planning is an integral process of order fulfillment (Seitz, et al.,2020, 1), see Figure 4. The objective of allocation planning is to fulfill customer order promises and to distribute the supply according to certain criteria. As well Kloos & Pibernik (2020, 203) highlight the importance of allocation planning to the comprehensive of order fulfillment management.

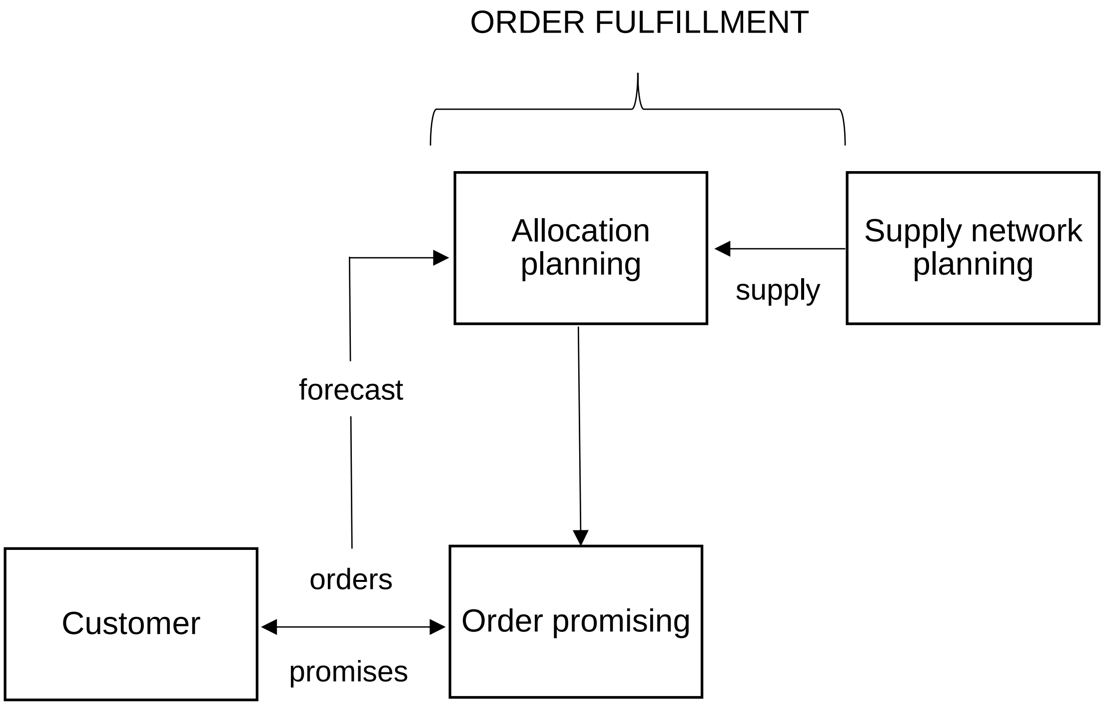


Figure 4. Order fulfillment process. Adapted from Seitz et al. (2020, 2).

Allocation planning is done based on existing customer demands. When supply is scarce and not all customer orders can be fulfilled, it comes to the decision to either strategically allocate the stock under service level agreement constraints or to adjust the inventory level (Thomas, 2005, 80). For a make-to-order manufacturing company where inventory is not held, first option of supply allocation according to certain criteria comes in question. Achieving high performance in terms of delivery reliability, steers the manufacturing companies into reaching improving performance. Delivery reliability is an important and integral aspect of order fulfillment processes. It is considered as competitive factor and there is high pressure in terms of performance excellence (Sridharan & Li, 2008, 1201).

Modern software systems such as advanced planning systems (APS) support allocation planning by incorporating models, algorithms, and operations (Stadtler, 2005, 575). Allocation planning is integrated for the software systems. Its objective is to allocate and reserve goods to from the inventory and incoming stock to certain customer groups. (Seitz, et al., 2020, 1). According to Jonsson et al. (2007, 832) APS is at help when managing today's complex supply issues and integrating the supply chain operations. They demonstrate several benefits of APS utilization such as cost optimization for merged companies and how APS can be integral tool while integrating marketing and production operations.

2.1.2 Manufacturing models

Make-to-order (MTO) system for manufacturing is a strategical decision of how production of the company is arranged and implemented. This means, production is initiated only after receiving customer order. MTO is the standard way of manufacturing for many industries, for example retail goods are usually manufactured in MTO basis (Bertsimas & Paschalidis, 2001, 119-120). As well automotive industry serves as a good example where make-to-order (MTO) manufacturing in widely in use and in addition industries such as restaurant business and furniture business also operate MTO basis (Xiao et al., 2014, 23). Make-to-order manufacturing is becoming more and more the norm for production method, as it can better serve the customer's diverse, special requirement needs in an industry, where products are non-standardized and non-customized. (Xiao et al., 2014, 30).

It is characteristic for MTO manufacturers not to hold inventories and the benefit of this system strives from the zero to minimized inventory costs. Controversially, the risk pursues from the company having to respect and go by production times i.e., lead times for their customers. (Og˘uz et al., 2010, 200-201). Any deviation of not following the lead time and the agreed service-level, might lead to consequences (Kloos & Pibernik, 2020, 204). These scenarios were discussed more in-depth in the service-level agreements chapter. In a case where manufacturer is unable to forecast the demand short-or-long term and/or has unique customized products for its customers, the likeliness of operating based on MTO is high (Sahin & Robinson, 2005, 538).

While some companies apply complete make-to-order (MTO) system, some companies have both make-to-order (MTO) manufacturing and make-to-stock (MTS) manufacturing systems in place. As well, there is several firms that operate between the systems, utilizing pros of both. (Rajagopalan, 2002, 241). Difference between MTO and MTS is that where MTO manufacturing is starting only after customer order is received, MTS manufacturing supports the customer demands being fulfilled from the existing finished goods inventory. MTS is especially used when purchased goods are universal, standardized, and high-demand products. MTS works based on producing the finished goods stock based on forecasting (Okongwu et al., 2012, 582). For example, company may apply a strategy, where low-demand orders are fulfilled make-to-order basis and high-demand orders fulfilled from make-to-stock finished goods inventory. (Hadj Youssef et al., 2004, 103-104).

Allocating orders in the make-to-order manufacturing systems is conducted by for example the first-in-first-out (FIFO) scheduling rules. This means, oldest released orders in the system have the allocation of the incoming MTO goods, arriving on stock. Company may also allocate the production capacity differently, by fulfilling small- and high-quantity customer demands in different order (Hadj Youssef et al., 2004, 104). The strategical decision where the allocation rules base on is considered in the order fulfillment processes.

Order promising is essentially a part of order fulfillment processes, where before the order is received, supplier commits to certain terms such as lead time and pricing to its customer. Aim of this process is to give as accurate information and to make promise of order fulfillment. (Venkatadri, et al. 2021, 2). Lead time covers time from order receiving to order delivery (Amer, et al., (2010, 282). For make-to-order companies the process of order acceptance and confirmation is critical. This comes in question when not all orders on hand can be fulfilled according to the customer requirements i.e., higher demand peaks and manufacturing and operations bottlenecks. The supplying company can choose between postponing the date, when lead time is longer than promised - order promise is not fulfilled - or delivering the order later than the confirmed date - negative impact on the delivery performance. First option may result in loss of revenue, but second choice may harm the supplier-buyer relationship or convert into penalties in case service-level agreement is applied. (Sawik, 2009, 6205).

2.2 Assessing delivery reliability

Delivery reliability is the second main concept under further investigation in this study, so it is worth highlighting the importance to both parties, particularly to purchasing company and explaining more in depth the concept and meaning. Delivery reliability reflects how the delivery from supplier to purchasing company is achieved against the promised order delivery date (Durugbo et al., 2014, 646). According to Sridharan & Li (2008, 1201) delivery reliability demonstrates the ability of the supplier to deliver parts on time.

As well for the manufacturers operating on make-to-order basis, enhancing delivery reliability is considered as important aspect. Timely diagnosis and review of performance are vital to understand actual delivery reliability. In case not sufficient performance is achieved, it should be investigated what the root cause for the later than promised deliveries is and how to eliminate its occurrence. (Soepenberg et al., 2012, 5491). Durugbo et al. (2014, 646) suggest that well working information flow regarding delivery reliability creates sustainable relationships and even improves the processes how delivery reliability measures are strategically formulated. Next chapters address the concept and importance of delivery reliability in the purchasing company perspective.

The SCOR-model invented by Supply Chain Council (SCC), is a comprehensive model for strategic supply chain decision making. SCOR is short for supply chain operations reference. Common issue for measuring supplier performance is that the metrics used, are different at each level of the supply chain. SCOR model is considered as alignment to tackle this issue. Operations focused SCOR model provides three levels of details, where the level one focuses on the performance metrics, divided into four categories: delivery reliability, flexibility and responsiveness, cost, and assets. (Huan et al., 2004, 24).

In short, the definition by Cirtita & Glaser-Segura (2012, 304) explains that delivery reliability is "*the ability to meet promised delivery date defined as on-time and in full shipments*". To scatter the supply chain delivery reliability into metrics, Figure 5 is created. Based description of the SCOR model, delivery reliability is one of the four supply chain performance attributes together with flexibility and responsiveness, cost, and assets. Delivery reliability comprises of a delivery of correct product, at the correct time, in the correct form in terms of packaging and condition, with the correct documentation provided and to the correct customer. (Cirtita & Glaser-Segura, 2012, 301).

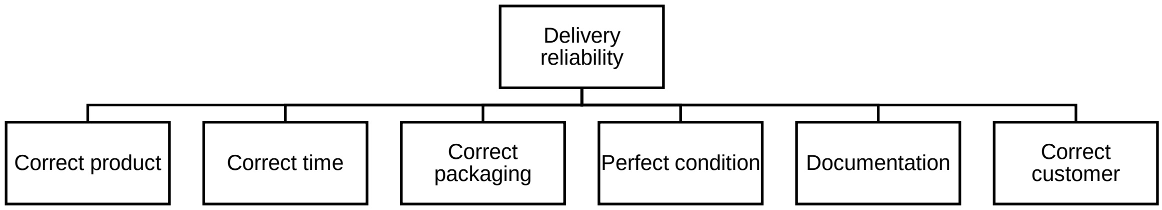


Figure 5. SCOR model for delivery reliability (Cirtita & Glaser-Segura, 2012, 301).

2.2.1 Measuring supplier performance

Supplier performance measurement systems is a "*set of metrics, used to quantify the efficiency and effectiveness of supplier's actions*" (Maestrini et al, 2018, 298). Performance measurement supports company's supply strategy and is a vital tool for strategical decision making. Performance assessment takes into consideration company internal measures and external measures, division is presented in Figure 6. When evaluating the supply chain operations measures, operations such as manufacturing, delivery process and reverse logistics are considered. External supply chain measurements are intended to measure relationships and operations external of company. (Maestrini, et al., 2017a, 300-302).

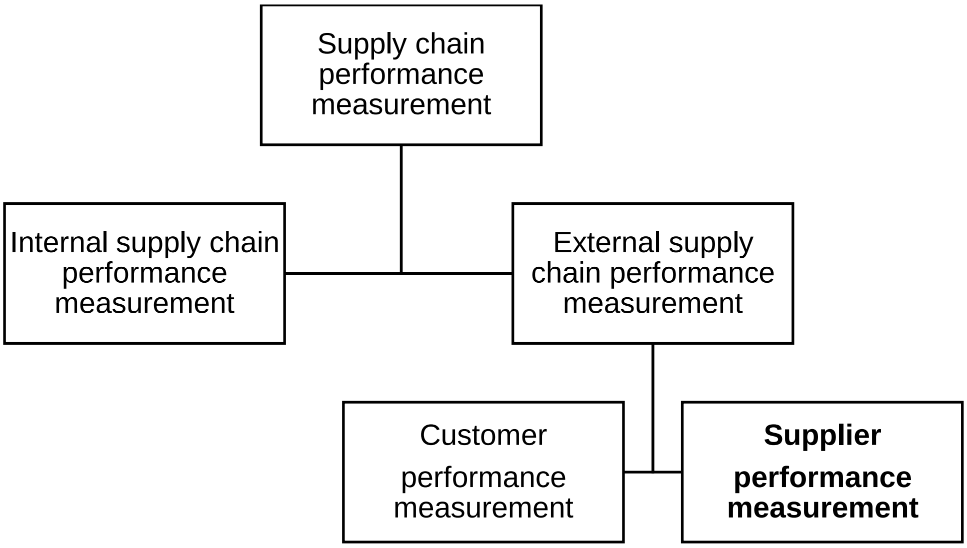


Figure 6. Supplier performance measurement system. Adapted from Maestrini et al. (2017a, 302).

Key performance indicators (KPI's) when measuring supplier performance are related to quality, delivery, and price. Measures for supplier performance give an overview of supplier's capabilities in terms of performance to the purchasing organization (Bourne & Mills, 2000, 765). According to Romule et al. (2019, 823) there is no universal standardized list of key performance indicators that should be used for supplier performance measurement, therefore they courage the selection of KPI's to be supplier specific. They present in their study five valued key performance indicators, used for supplier performance measurement purposes. These are,

- Net profits, including supplier development activities return of invest, supplier profitability, economic value, and cost of quality.
- Delivery performance and time, that considers performance measures on delivery, short- and long-term financials, and logistics.
- Flexibility and responsiveness comprise of activities on supply chain capabilities, strategy formation and supplier flexibility and monitoring.
- Product quality and availability are quality management and assessment related activities. These include product availability.
- Cycle time means for example strategical supplier selection. (Romule et al., 2019, 821-825).

Relationship factors such as commitment and timely communication are important measurements in buyer-supplier relationships. These types of non-price related factors have positive relevance to growth in sales, customer service and competitive position of a company. (Simpson et al., 2002, 33-38). Romule et al. (2019, 831) agree in their study when emphasizing the flexibility and responsiveness being one of the most important supplier performance KPI's. They as well highlight the importance of the supplier's perspective and argue that it should be included when setting and implementing the measurements. According to them, supplier assessment derives into performance measures. Delivery reliability performance and time are considered as integral key performance indicators. (Romule et al., 2019, 831)

The life cycle of performance measurement systems (PMS), in Figure 7, consists of use, design, implementation, and systematic review phases. Next chapters focus in emphasizing the most relevant phases of design, use and review to this study context.

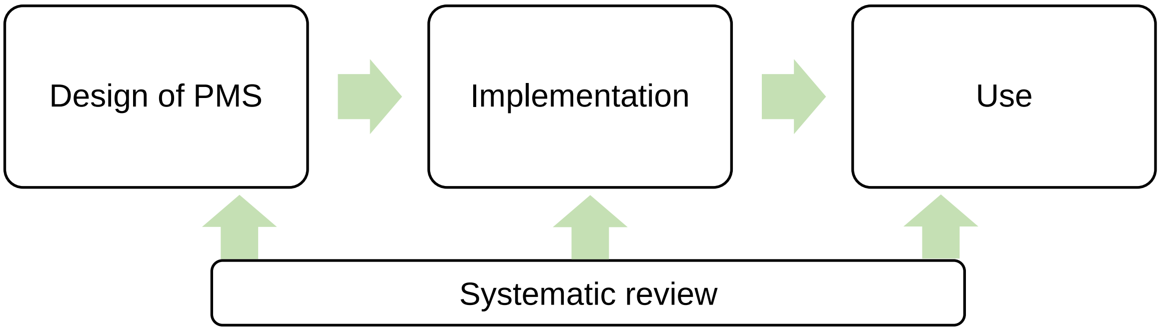


Figure 7. Life cycle of performance measurement system (Maestrini et al., 2017b, 2042).

Design

Starting point for the performance measurement system is the design phase. Two steps are completed during this phase: identifying and deciding on key performance metrics and designing the measures i.e., target performance to those. (Bourne & Mills, 2000, 758). Alignment with the company overall strategy and supply strategy is vital (Franco-Santos, et al., 2012, 79). According to Melnyk et al. (2014, 175) it is important to set the measures correct way. When setting up measures, there is three aspects to consider, (1) select measure that reflects to the best extent, what is happening. (2) standardize target-level to see how supplier performance reflects and compares to the expectation. By comparing the quantified performance to the standard performance level, company can identify weather the supplier performance is on a wanted level. (3) determine the consequence of falling below or reaching above of the standard target-level. (Melnyk et al., 2014, 175).

The key to successful design phase is to consider all aspects from company overall strategy and supply strategy. These strategies reflect the metrics that are set during this life-cycle phase. (Franco-Santos, et al., 2012, 79–80). Maestrini et al., (2018, 302-303) study demonstrates that suppliers should be included in the design phase, in order to ensure the selection of right metrics and customize the performance measurement to match the partnership flow of the supplier in question. What is comes to the other stakeholders, it is important to include those as well in the design phase of performance measurement systems. Poorly functioning performance measurement system is usually already detected at this stage, and the most common reason behind the malfunctioning performance measurement system is company not being able to limit the performance metrics properly. Limitation and delimitation are vital for the success of the next life cycle phases. Existing challenge is that there is no universal performance measurement system fit for all purposes and all buyer-supplier relationships. (Gopal & Thakkar, 2011, 526–529; Maestrini, et al., 2017b, 2042–2043).

Implement

The essential of implementation phase is to establish the performance measurement system on a selected platform by utilizing supporting systems. To complete this phase, a working IT infrastructure is necessary. The objective of implementation phase is to have the processes and systems ready for supplier performance measurement activities. According to Maestrini et al., (2017b, 2043) incorrectly performing data reporting tools may provide false figures that mean unreliable performance measures. If false measures are reported to the supplier, it may create distrust between the buying company and supplier. Reliable data ensures long-term supplier performance measurement in a standardized way. (Bourne & Mills, 2000, 758; Maestrini, et al., 2017, 2043).

Use

Use phase encovers how well the implementation phase is completed. (Bourne & Mills, 2000, 758). Activities for the use phase are for example communication and feedback giving and managing incentives and penalties. First activity concerns the reporting of performance toward supplier and providing possible improvement action plan to enhance the performance. Latter activity grants incentives for good performance and claims penalties in case of falling back from the expected performance levels. The SPMS use may be approached by two ways, diagnostic or interactive way. In first approach buying company controls and reports the supplier performance via the SPMS's. When targets are reached, buyer company grants incentives and when the supplier is falling behind the target, buyer company claims penalties. The latter interactive approach emphasizes cooperation, open discussion, transparency in sharing the performance results and drives for mutual benefit. Both parties have mutual goals and motivation to improve the performance in collaboration. (Maestrini et al., 2017b, 2043–2044). Henri (2006, 531) highlights that both approaches can co-exist and complement each-other, with the goal on improving and shaping the SPMS even further.

Study of Maestrini et al. (2018) propose that suppliers adapt indifferent approach to the feedback received or demonstrate passive or active interest in the feedback. When approach to feedback is indifferent, supplier takes no interest regarding the measurement. This may derive from the positioning of the supplier in a supplier-dominant market, see Kraljic's model in Figure 8. For the passive approach it is characteristic that supplier takes the review into consideration but does not interactively respond to the feedback giver. Active interest showing suppliers actively and transparently communicate their thoughts to the evaluation giver. This may derive from the supplier's status as strategic supplier and by adapting partnership type of relationships. (Maestrini et al., 2018, 304). Prahinski & Benton (2004, 60) study shows that collaborative communication such as feedback giving regarding developable objectives is perceived to improve supplier-buyer relationships. They note that buying company can impact on supplier's willingness to commit to the relationship by communication and relationship commitment. (Prahinski & Benton, 2004, 60).

Review

Systematic review of performance measurement system enables an up-to-date processes and measures (Bourne & Mills, 2000, 758-759). Especially in terms of supplier performance measurement, as inter-company process, it is important due to the complex supply flow structure. The review should be conducted on regular basis and by systematic approach. Changes intra organizationally and environmentally steer the measurement systems. (Melnyk et al., 2014, 175). In the long run, performance measures can naturally develop to the right direction. It is wort noting that if measurement system is left without attention and systematic review, company may end up with an old, non-up-to date system. Measurements would be no longer valid in the current company context, nor supporting the company strategy. Then the complete process would have to be created from scratch (Bourne & Mills, 2000, 768).

Findings of Maestrini et al. (2018) study show, that supplier performance measurement in fact has a positive impact on the evolvement of supplier's performance: the phases of use and review have the greatest impact on the supplier performance. It is not sufficient to evaluate how supplier has performed against what was contracted now. Supplier performance measurement evaluation should be long-term and considering future improvement aspects (Patrucco et al., 2021, 8). Patrucco et al. (2021, 9) study shows that supplier performance measurement has an impact on supplier's commitment. By ongoing evaluation buyer demonstrates active focus in supplier's activities and on the relationship. Although buying companies aim for financial benefits with setting up performance measurement systems, there is a lack of evidence that it turns into financial performance for the company (Koufteros et al., 2014, 331). Focus for the managers should be in the proactive approach to using supplier performance measurement systems and review and adapting the performance indicators in the changing business (Maestrini et al., 2018, 2056).

2.2.2 Supplier evaluation and development

Performance of the supplier affects widely supplier relationship management (SRM) practices. In supplier evaluation practices, performance of the supplier in terms of quality, cost and delivery is considered as criteria for evaluation, together with supplier capabilities and collaborative relationships. To some extent, supplier selection and evaluation are integrated under the supplier relationship management (SRM) practices. Baskaran et al. (2012, 648) summarize the connection well by arguing that "*supplier evaluation is a management* *decision-making process that addresses how organizations select strategic suppliers to enhance their competitive advantage*”**.** For supplier development practices, the main objective is to enhance the supplier performance. (Park et al., 2010, 506).

While outsourcing is increasing, it is the aim of correct supplier choice to ensure quality standards being fulfilled, optimal cost and timely deliveries to the company (Yang et al., 2017, 105). Correct selection of supplier is not a simple task but with comprehensive supplier evaluation activities the aim of selecting well performing supplier can be achieved to best extent. (Huang and Keskar, 2007, 510-511). Main purpose of supplier evaluation is to first identify correct metrics to evaluate suppliers with and then to measure and monitor supplier performance. (Park et al., 2010, 501-502). Well performing suppliers are seen as connected to the success of entire supply chain (Kannan, 2018, 392).

Supplier development activities come in question when buyer is dissatisfied with their supplier who is falling behind on the performance expectation levels. In such circumstances, buying company has two strategical ways to steer towards: either to discard the mal-performing supplier and find alternative source to buy from, or to engage into supplier development activities with the aim to improve the performance of the current supplier. Supplier development is important as buying company may not always have the first option to be selected from. Supply may be scarce without alternative suppliers existing in the market. (Krause & Ellram, 1997, 21-22).

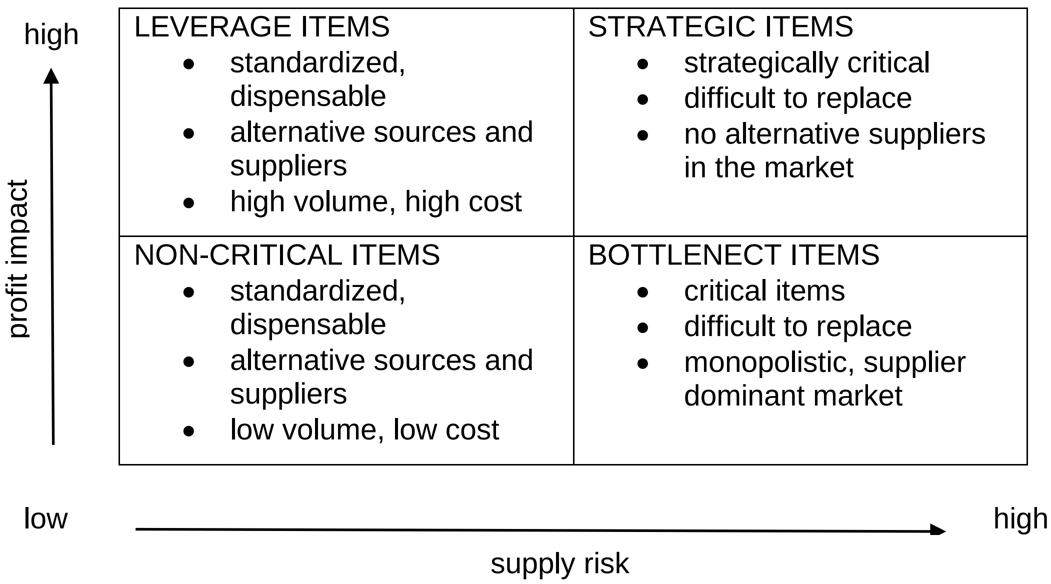


Figure 8. Kraljic’s purchasing portfolio matrix. Adapted from Park et. al (2010, 500).

Using the first option of switching the supplier is depending on the product placement in the product portfolio, presented in Figure 8. Kraljic's model is shown here to emphasize the importance of supplier development actions and to underlie realistically that there might not be any alternative choices than to work with and try to improve poorly performing supplier performance. As mentioned, it may be that there is no alternative supplier to source from, in case of a supplier-dominant market situation. (Cees et al., 2005, 20-21). Then again, for the non-critical and leverage products, changing the supplier with the hope to receive better performance from alternative source, should be obvious in order to secure the supply and incoming goods. (Friedl and Wagner, 2012, 3066-3068). The portfolio matrix in Figure 8 categorized the supply risk into low and high risks. When supply risk is considered high, i.e., for strategic and bottleneck products, a cooperative strategy is to be adapted as there is much of relationship relevance involved in supply flow. Controversially, when supply risk is low, competitive strategy takes place which emphasizes the power dominance being at purchaser's side in forms of for example cost reduction. (Park et al., 2010, 500).

According to Krause & Ellram (1997, 21) supplier development is considered to be any sort of effort to enhance the supplier performance. Targeted improvement objects are for example quality topics in terms of products and services, delivery reliability and delivery accuracy, cost reduction and lastly new technology adoption and utilization. Supplier development actions approach can be either reactive or proactive at the purchasing company. Companies approaching supplier development reactively, act only when there is already an existing risk and chance that company's internal processes are harmed due to lacking performance of the supplier. Risks such as production standstill or internal complaints might have already occurred. Companies utilizing reactive approach in supplier development consider supplier development as beneficial for both parties in terms of maximizing profit for both. Actions are taken prior any internal complaints and supplier performance is monitored on a regular basis. (Krause et al., 1998, 40-45).

Systematic approach with the target to support supplier development activities is referred to as supplier development programs. The main objective of such programs is to support the performance increase of supplier short and long-term. Khan & Nicholson (2014) present three main supplier development activities that can be adapted: direct investment of capital, acquisition of the supplier company, even partially and human and organizational investments. Last one emphasizes the role of knowledge transfer and sharing know-how information. As well the personnel involvment in supplier development programs is emphasized. Top management involvement, supplier production plant visits, knowledge transfer and training of personnel all serve as good examples of supplier development investment activities. (Khan & Nicholson, 2014, 1212-1215; Arroyo-López et al., 2012, 681-684).

2.2.3 Risks affecting delivery reliability

Acknowledging the concerns of risks in the context of supply chain management has gained great emphasis among researchers and company management. Systematic planning and approach to how to mitigate disruption risk is an essential element of company's overall strategy (Tomlin, 2006, 655). To start with the definition of supply chain risk management, Sodhi et al. (2012,1) research elaborates that, SCRM is according to majority of their study respondents, about managing supply-chain randomness and unexpectedness. Also, it is about dealing with supply chain related operations and handling unknown disruptions and financial risks. (Sodhi et al., 2012, 1, 7). Other definition is provided by Nooraie & Mellat Parast (2015, 193) where supply chain risks "*can be regarded as risks associated with incidents such as an unanticipated event within a supply chain and the associated negative outcomes of that event on the supply chain*". Uncertainty is recognized element of today's business. As competition is getting more intense, companies are forced to form optimized supply chains where the decentralized formation makes the entire chain exposed to unknown threads. (Baghalian et al., 2013 199).

Researchers in the supply chain risk management (SCRM) field categorize supply chain related risks various ways. For example, Chopra and Sodhi (2004) consider risks occurring as disruptions or delays and risks in supply chain relate to system use, forecasting, intellectual property, receivables, inventory, and capacity. According to Tang & Tomlin (2008, 12-13) there are six major supply chain risks, that take place frequently in the supply chains. These are: supply risks, process risk, demand risk, intellectual property risk, behavioral risk, and political and social risks (Tang & Tomlin, 2008, 13-14). Diabat et al. (2012, 3039) provide to some extent narrower categorization of risks relevant to supply chains. This categorization consists of supply risks, operations risks, and demand risks. Next the risks of this categorization are elaborated more in-depth. Supply risk and operational risks are evaluated as they are important for the purchasing company whereas demand risk being important for the supplying company.

Supply risks include movement of all materials in the supply chain – from supplier to manufacturer, reliability of suppliers, decision of single-source use in comparison to multiple-source use and decision of centralized and decentralized purchasing. (Diabat et al., 2012, 3039). Tang & Tomlin (2008, 13-14) consider supply risk via two risk types. First one being the supply cost risk, which may cause extra costs for the purchasing company due to increasing material costs at suppliers. Second type is the supply commitment risk where again purchaser is to some extent committed to the supplier via extensive contracts and ordering rules. (Tang & Tomlin, 2008, 13-14).

Operations risks consist of company's internal capabilities in manufacturing, profitability risks, process risks and technology risks (Diabat et al., 2012, 3039). Company's internal operations are considered as operational risks and therefore company's inbound and outbound logistics operations are a source of risks and may impact company's effectiveness, capabilities, and quality requirements (Tang & Tomlin, 2008, 13). For example, delays or capacity issues at operational outbound logistics may be the root cause in delay of customer deliveries.

Demand risks often are associated to demand fluctuations. Demand risks result from the goods flow from company to customer including risks related to inventory planning in form of stock-outs and excessive stock (Diabat et al., 2012, 3039). Agreed by Sodhi (2005, 72) who's study considers demand risk measures two ways, either not fulfilling customer's demand or by having excess inventory. Demand risks may be associated in operating in certain countries. As well the volatility of demand is considered a risk for the supplying company operations. (Tang & Tomlin, 2008, 13).

Tang & Tomiln (2008, 14) introduce two measures to consider when assessing risks. First one concerns assessment of the likeliness of the risk to occur. Second evaluates the negative impact of risk occurrence. Risks are of nature undesirable and usually impact negatively on the company operations and profit. Supply chain risk management strategies focus on mitigating both measurable risk types by downsizing the likelihood of risk occurring and by reducing possible negative impact affected by the occurred risk. (Tang & Tomiln, 2008, 14).

According to Tomlin (2006, 640) companies tend to approach risk management either by mitigating the risks in advance or acting at the time of occurrence. For the first tactic the company carries the cost of preparing for the risk, even so the risk occurrence may never take place. The second contingency tactic is reliant on the supplier's capabilities. It can be that companies adopt both tactics. (Tomlin, 2006, 640). According to Nooraie & Mellat Parast (2015, 197) supply chain visibility can be viewed as supply chain risk mitigation strategy as it significantly reduces costs at the event of risk occurrence. This applies for both supply and demand risks. By supply chain visibility it is emphasized that information sharing amongst the supply chain partners regarding customer demands, inventory and transportation costs increases power and decreases uncertainties. (Christopher. & Lee, 2004, 391).

3 METHODOLOGY

This chapter presents the methodological approach of this thesis. Aim is to walk the reader through the stages by which the empirical analysis was conducted. First, the research design is presented to get the complete overview of the methodological choices. Second chapter elaborates the process of the data collection including choices for the sampling of the data. After that, the statistical analyses used for analyzing the collected data are justified, together with the presentation of hypotheses setting. Methodology chapter ends by comprehensively and to best extent, objectively assessing the reliability and validity of this research.

3.1 Research design

This research is a case study conducted in one organization setting. A single-case study approach is selected for this thesis due to its manageability and suitability to have deeper insights of one organization setting. According to Voss et al., (2002, 195) a case study is especially good when wanting to establish new insights and when developing a new theory. In general, a case study focuses on a real-life phenomenon or topic. The selected case may be a person, group, organization, union, or an industry, to name a few. What is important in case studies is the selection of the case and making limitations to it. Case-study research allows having in-depth insights of the phenomena. Case research is especially suitable for operations management study context (Voss et al., 2002, 195). Criticism towards case studies concerns the research insights not being generalizable and reliable enough. (Saunders et al., 2016, 184-186). Agreed by Voss et al. (2002, 195) who mention that one of the challenges in case research method is producing generalizable findings.

The research is carried out by using quantitative methods. Approach for this research is a statistical analysis with the aim to test hypotheses. Statistical analyses are carried out by using Stata software, a program for statistical analysis purposes. Research is conducted as archival and documentary research, also referred as desk research. Where the secondary data is gathered from company internal databases, websites, and archives. The sources used for this research are reports – that are in the first place generated for performance evaluation purposes, contracts, and agreements. Quantitative research is characterized with numeric data, when compared to qualitative research where the data is in non-numeric form, for example in words or pictures. As in this research, quantitative research often identifies via highly structured data collection methods. (Saunders et al., 2016, 165-166).

Main objective when using quantitative research methodology is to investigate and interpret relationships between variables, when the variables are measured numerically (Saunders et al., 2016, 165-166). The quantitative data of this research is presented as numerical data, where values of variables are measured in numerical values (Saunders et al., 2016, 500). Numerical measures are precise and leave not at all to very little room for own interpretations. In below Figure 9 the research design choices are collected.

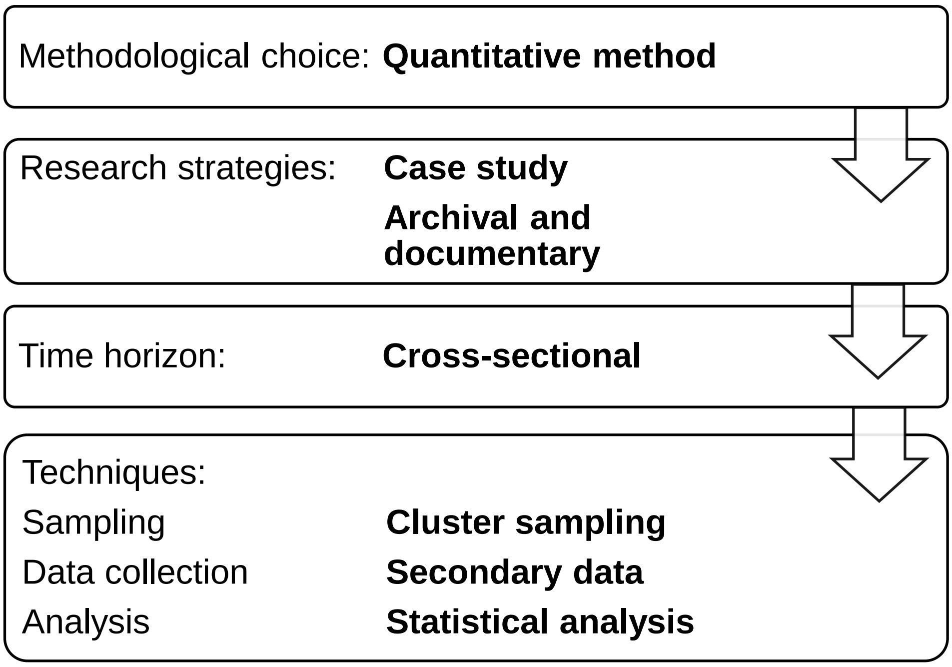


Figure 9. Reseach design steps. Adapted from content of Saunders et al. (2016).

The intention in this research is to test possible linkages between service-level agreements and delivery reliability. As well, to find out where the possible fall below the perfect delivery reliability is deriving from and to evaluate how case company is performing as supplier in terms of delivery reliability. The aim of this research is to provide in-depth analysis via quantitative methods and answer research questions set at the beginning of this thesis.

3.2 Data collection

Empirical analyses of this thesis are carried out by utilizing secondary data sources. According to Saunders at al. (2016, 318-319) secondary data covers three different data settings: document-based data, such as company databases. Survey based data such as indexes and attitude surveys. And multiple-source data in case of for example big data sets and industry wide reports. The data for this thesis is document-based data, collected from case company's internal databases. The data collection process was conducted in two stages, see Figure 10. Stage one was the formulation of raw data, which according to Saunders et al. (2016, 3018) is only to some extent or not at all processed data from the company internal reporting and analyzing tool. As an output, a dataset presenting actual delivery reliability figures from 2019-2020 per customer and customer group was derived. In stage two, company internal archives and data-storing folders were searched to find current customer contracts. The contents of these contracts and agreements were evaluated to identify the characteristic components of service-level agreements and to take decision whether the contract was sufficient to be counted as SLA for this research. Stage two evaluation outcome was a vital preceding step for the later conducted sampling phase.

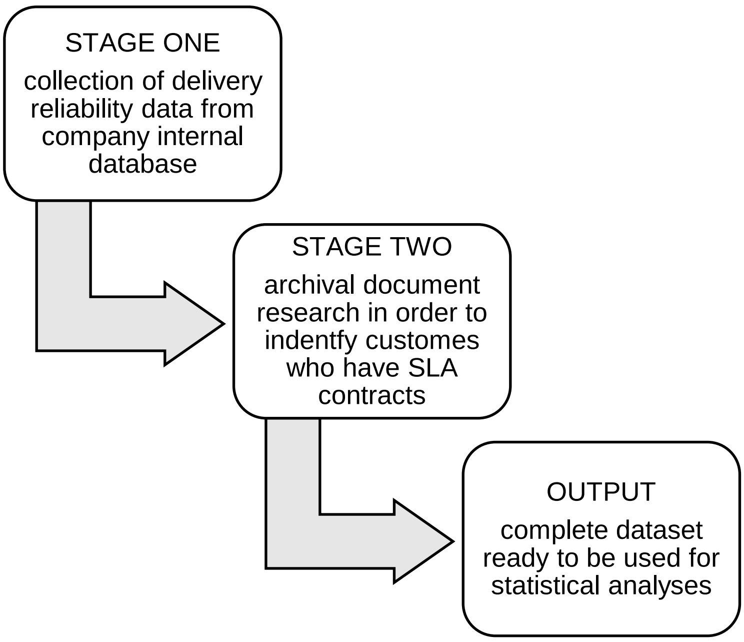


Figure 10. Data collection stages.

Time horizon for the research was selected to be two years, 2019-2020 and the observations are on monthly level, totaling the possibility of 24 observations per customer. With this long review period the aim is to minimize any COVID-19 crisis impacted delivery disturbances. As well, any market impacts such as global component and raw material shortages that usually have a negative impact on the timely ability of finished goods in spare part business. Another justification was to have enough analyzable data to conduct the statistical analysis with. Aim was to have the data to describe as normal as possible situation for case company.

Stage one report is in the form of long panel data matrix. Based on the stage two archival data collection, companies in the stage one report are divided into two groups, depending on if they have SLA in use or not. SLA companies have had the SLA in use already at the beginning of the observation horizon, prior to January 2019, this is notified during the stage two data collection. Another criterion the customer had to fulfil in order to be classified as SLA-customer (Group 1) was the characteristics of the contract. Firstly, customer agreement regarding delivery reliability, logistics or service level had to be found on the company internal archives. Secondly, the contract had to consist of majority of the SLA components (service-level expectation, requirement, review horizon, consequence of falling below expectation and so on). Third, the contract had to be signed by both parties, case company as supplier and customer as buyer. Total of 1070 companies are included in the report. From these, 83 companies are defined as SLA companies and 987 are defined as non-SLA companies.

It is worth mentioning regarding the stage two archival data collection, that all findings of the statistical analyses that are reflected to customer expectations set in the contracts, are expressed in very general level in the next chapters. As the stage two data collection concerned the content of the contracts, the nature of business contracts is strictly confidential, and therefore, for example consequences of failing to meet delivery reliability expectations are not at all covered. In general, it can be professed that majority of SLA customers utilized a review horizon of one month and a delivery reliability percent between 80 to 100 %. This information in needed in later phases, when evaluating the performance of the case company and responding to the research questions. Latter information regarding the delivery reliability expectation is especially relevant, as it supports answering the research questions by indicating what a proportion of customers are expecting in terms of delivery reliability. It is seen as essential information when evaluating at what level case company is expected to perform by the customers.

Concerning the delivery reliability figures in the data matrix, three figures to measure delivery reliability were calculated in the data for analysis purposes. Overall delivery reliability measure per month per customer consists of three variables. First variable is the number of missed on-time deliveries caused by supplier delay (supplier hits). Second variable presents the number of missed on-time deliveries caused by warehouse delay (warehouse hits). Third variable is the total numbers of deliveries that took place during the observation month (order lines total). Thus, the actual delivery reliability in percent was calculated with following formula:

$$\text{delivery reliability} = \frac{\text{order lines total - (supplier hits + warehouse hits)}}{\text{order lines total}} \text{ x 100 (\%)}$$

For the better suitability in terms of analysis purposes, delivery reliability figures (delivery reliability, supplier delivery reliability, warehouse delivery reliability) are left as decimal number. The delivery reliability measure and the scattered components of it, give a total of seven variables to be utilized in statistical analyses. Below Table 4 is a miniature model of the long panel data matrix dataset. Maximum number of observations per customer is 24.

Table 4. Long panel data matrix.

Obs	Customer	Group	MM. YYYY	Order lines total	Missed delive-reies total	Supp-lier hits	Supp-lier deli-very re-liability	Ware-house hits	Ware-house deli-very re-liability	Deli-very re-liability
1	Customer 1	Group 1	01.2019	88	13	13	0,852		1,000	0,852
2	Customer 1	Group 1	02.2019	92	14	14	0,848		1,000	0,848
3	Customer 1	Group 1	03.2019	136	11	11	0,919		1,000	0,919
4	Customer 1	Group 1	04.2019	101	4	4	0,960		1,000	0,960
538	Customer 31	Group 2	01.2019	161	63	63	0,609		1,000	0,609
539	Customer 31	Group 2	02.2019	134	35	35	0,739		1,000	0,739
540	Customer 31	Group 2	03.2019	165	32	32	0,806		1,000	0,806
541	Customer 31	Group 2	04.2019	115	65	63	0,452	2	0,983	0,435

It is important to understand what happens in case of missing data in the data matrix. In a survey-case setting, this could be a non-response by a survey respondent. (Saunders et al., 2016, 507-508). In desk research setting, missing data means a missing value from any observation month. There are few reasons why in this research context data is missing. First is that customer deliveries have started after the data collection start period, meaning after first observation month of January 2019. Second point is the contrary, that customer has exited the business during the observation period of 2019-2020. Third, is that there might have been zero deliveries to customer during an observation month, for example in a case of customer's holiday closure. Luckily, the format of data being in long panel data form, supports the missing data. It simply does not have any value for those months, where there were no deliveries i.e., no data available. In theory, all customers selected for the sample groups should have 24 observations, but some, for above reasons, may have less than that. This is especially convenient when evaluating the impact of outliers in data setting.

3.2.1 Sampling

Sampling is needed as analyzing the whole population is inefficient and time consuming (Saunders et al., 2016, 277). Same idea is applicable in the context of this research. As mentioned, while discussing the data collection, a total of 1070 companies are included in the raw data of delivery reliability report. It is impossible to expect the customers at this wide range to be comparable. Sample groups intention is to give as good as possible picture of the population.

Sampling technique selected for this research is cluster sampling and by that two customer groups are formed, based on if they have SLA (Group 1) or no SLA (Group 2). Cluster sampling is particularly used when the sample criteria is for example a type of company or geographical location (Saunders et al. 2016,.291). There are few criteria the sample size is formulated for later analyses purposes. First, the customers selected for the sample need to be located in Europe (geographical criteria). Second, customers selected for the sample are the top 30 biggest customers in terms of number of order lines on average, during the review horizon of 2019-2020 (size criteria). The average is calculated by sum of the observations divided by the number of observations. This criterion is screened as the first step when forming the two sample groups. Saunders et al. (2016, 280) encourage the sample size to be more than 30 cases for statistical analysis purposes. This research follows this recommendation by formulating two sample groups based on selecting the biggest customers in terms of the average no. of deliveries per month.

By looking at the data collected and formulated via the stage one and stage two data collection, it is obvious to state that there are enormously more of Group 2 companies (those who do not have SLA) than Group 1 companies (those who have SLA). As table 5 shows, the percent in the raw data of Group 1 companies is 7,76 % when compared to the Group 2 companies that totals at 92,24 %.

Table 5. Customers in the raw data prior to sampling.

Group	Definition	Total	Total (%)
1	SLA customer group	83	7,76 %
2	non-SLA customer group	987	92,24 %
	Total	1070	100,00 %

There were two possible approaches to select the sample companies according to the size criterion. First option was to select top 30 companies from both sample groups (Group 1 and Group 2), based on the average of delivered orders. This would mean the sample groups being equal in number of companies (50 % / 50 %). Second option was to decide on a level of minimum average (deliveries per month), in a way that Group 1 sample would reach the recommendable sample size of 30 cases. In this research the minimum number of orders on average needed would have been 7 or more order lines per month.

The essential question of the sampling phase was if the sample groups should be more equal in numbers rather than equal in customer size. The decision was eventually taken to discard the option two and to adopt option one: sample groups equal of size, meaning the top 30 companies in terms of average of order lines during the observation horizon. Sample groups, with min and max values for variable order lines total are presented in Table 6.

Table 6. Sample groups.

Group	Top 30 companies in terms of average of total order lines	Max. Value	Min. Value	No. Of Custo-mers	% of total
1	SLA customer group	681	8	30	50 %
2	non-SLA customer group	734	50	30	50 %

3.3 Analyses design and hypothesizing

As mentioned, the main objective of this research is to investigate possible linkages between service-level agreements and delivery reliability and to provide statistical analyses the research questions may be answered by. This is reached via hypothesizing and testing the hypotheses with statistical analyses. Conceptual framework of this thesis is presented in Figure 11. There were total of four hypotheses drawn for statistical analysis's purposes. Fifth object in the framework, named to refer to the sub-research question 2 (S-RQ2), was as well included in this framework, due to its relative importance when responding to the research question regarding case company performance as supplier. Answers to this sub-research question is as well seen in the statistical analyses but there was no actual hypothesis set for this object, as it is more evaluative in nature.

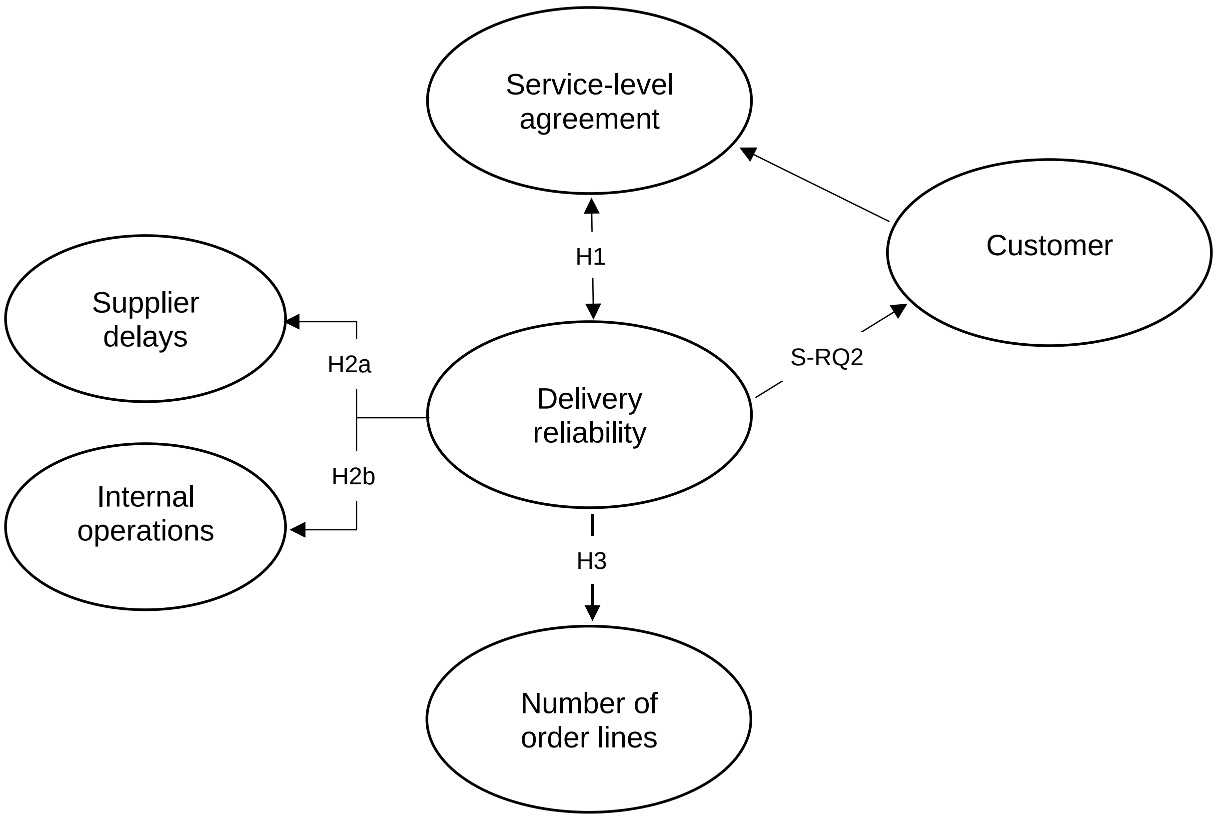


Figure 11. Conceptual framework.

To start with understanding the data set comprehensively, a descriptive statistics table is created to present the data best possible way. As well, the descriptive statistics supports in providing values of the data. Having that focus, for all the seven variables in the data matrix that are defined more in detail together with the descriptive table, a descriptive statistics table was formulated, that shows:

- no. of observations
- mean
- standard deviation
- minimum value
- maximum value

Number of observations tells how many observations were per each variable. Minimum and maximum value demonstrate the lowest and highest values of each variable. Mean is the average of variable results, calculated by adding all results together and dividing by the number of observations. Saunders et al. (2016, 531) define standard deviation as to "*how the values of the variable are spread around their mean*". In other words, the closer the values are to the mean, the more typical the mean is. On the contrary, if variables vary widely around the mean, mean is unlikely to be typical. (Saunders et al., 2016. 531).

3.3.1 T-test

For an efficient testing by statistical analyses, a t-test between two sample groups is chosen to be sufficient for analysis purposes. T-tests are commonly used for statistical analyses and business research due to its suitability for many situations. T-test compares statistically the means of two groups and show if the difference between these groups is statistically significant. When the aim is to test if two groups are different – as it is in this research – a t-test is suitable for such purposes. T-test is applicable when numerically measured variable can be divided into two separate groups and the aim is to assess the likelihood of these groups being different. (Saunders et al., 2016, 541-542).

Geher & Hall (2014, 235) note, that a t-test is suitable also for smaller sample sizes, as it is difficult to reject the null hypotheses. (Geher & Hall, 2014, 235, 248). As mentioned, in this research the data is in the form of long panel data matrix although showing results on monthly basis. The t-test is considered as suitable for this empirical research context, as the interest is not to detect causalities over time or trends in the market, but to have a statistical analysis of the actual delivery reliability figures over selected time span.

The t-test consists of two parts, where the first part of the test is to test for equal variances of variables. T-test assumes that variances are equal (Aljandali, 2016, 90), but Stata software used in this research for these analyses, supports testing of also non-equal variance variables. Second part is the statistical comparison of independent variable groups means. Result of both analyses steps is a probability, p-value scale, where p-value determines the level of statistical significance (Saunders et al. 2016, 544). P-value determines the confidence ratio, for example a confidence of 95 % that is applied in this research context. This means that the statistical significance can be noticed if result of the t-test is greater than 0.05 ($p < 0.05$). A result being greater than this, would lead to discarding the null hypothesis. (Infanger & Schmidt-Trucksäss, 2019, 4191). Null hypothesis is the statement of “no difference in means found”. Hypothesis is set as: H0: means of the two samples are same and HA: means of the two samples are not the same. The alternative hypothesis (HA) is accepted, if it is noted via statistical test that null hypotheses in unlikely to be true (Aljandali, 2016, 90).

Hypotheses for the first part of all three t-tests was formulated for all variables same way:

H0: Variances are equal
HA: Variances are not equal

Hypotheses for the second part of the tests were formulated as follows:

H0: Delivery reliability mean is the same for the sample groups
H1: Delivery reliability mean is not the same for the sample groups

H0: Supplier delivery reliability mean is the same for the sample groups
H2a: Supplier delivery reliability mean is not the same for the sample groups

H0: Warehouse delivery reliability mean is the same for the sample groups
H2b: Warehouse delivery reliability mean is not the same for the sample groups

3.3.2 Spearman’s correlation

Before selecting Spearman’s correlation test for non-normally distributed data, the non-normality needs to be proved. This is conducted by running a Shapiro-Wilk test for testing if data is normally distributed. The result of this test indicates which correlation test, Pearson or Spearman should be continued with. A probability value, p-value indicates if the data is normally distributed or not. Probability of 0.05 proposes that all values below this are non-normally distributed and values above, suggest for normal distribution. Spearman's correlation test is selected over Pearson's correlation in cases where data is non-normally distributed. (Saunders et al., 2016, 535-537). For the first part, testing for normal distribution, hypothesis was formulated as below:

H0: Variable is normally distributed
HA: Variable is not normally distributed

Spearman's correlation expresses the strength between two variables. Result of the Spearman's correlation is between -1 and 1. Where -1 indicates perfect negative correlation and 1 perfect positive correlation. The scale and level of strength is presented in below Figure 12. Result 0 means there is no correlation and the two variables tested are independent from each other. (Saunders et al., 2016, 545). When assessing the strength of the relationship of the two variables in this research context: order lines total and missed deliveries total, the Spearman's correlation analysis is executed with following hypothesis:

H0: There is no correlation between variables
H3: There is correlation between variables

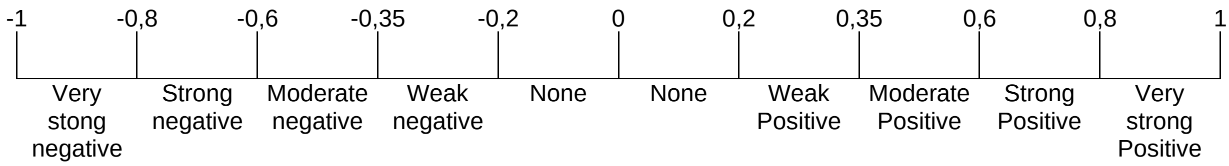


Figure 12. Strength of correlation. Adapted from Saunders et al. (2016, 545).

3.4 Research reliability and validity

While assessing reliability and validity of a research, it is important to maintain the objectivity to the best extent. Research reliability refers to the replicability of the research. It can be assessed with the question "what if the study was conducted again, would the findings be the same?". If the answer is yes, the study is considered to be reliable. (Saunders et al., 2016, 202). Therefore, when assessing the reliability of this research, it can be concluded that the reliability in on a good level. What supports this statement, especially when assessing the reliability of the data from company internal reporting tool, is that the numerical data regarding the actual delivery reliability is implicit. With quantitative methods there is very little room for writer's own interpretation, and for the dataset gathered to give false answers. Therefore, any, for example, participant biases and errors have not taken place in the dataset. As well, the statistical analyses were each carried out minimum two times, to minimize any typographical errors. With statistical analysis there is always the risk of false interpretation by the researcher, and this concern is minimized to the best extent by multiple time checking of the results. Any misinterpretation of the statistical analysis reports is ensured with thorough examination of the results and by studying the theory of statistical analyses comprehensively.

Naturally, this research relies on the company internal reporting tool from where the data is gathered to be reliable and giving correct results. As well, for secondary data it is worth considering why the data is generated in the first place and for what purpose. Although, the data used in this research is secondary, it is still intended for the exact same purpose in real-life than in this thesis: to assess the delivery reliability towards customers and to use that information for, for example KPI reporting purposes. In nature, the dataset qualifies as good and relevant as primary data would.

Then again, what comes to the stage two data collection form company internal archives, the reliability presumption needs closer attention. There the mission was to go through all documents regarding customer supply chain contracts, delivery reliability contracts, service-level agreements, logistics agreements and meeting minutes to identify customers that classify as service-level agreement customers (Group 1). What was noticed during this data collection phase, is that the contracts and agreements varied widely in quality. The main challenge was to identify and take the decision on the documents if they would be classified in this research to have enough characteristics to be evaluated as service-level agreements. For the research replicability, it was anyhow essential to make the sampling limitations. Eventually, a decision was made that to be classified as an SLA customer (Group 1), the document in question had to include at least come SLA characteristics (i.e., service-level expectation, review horizon, consequence) and be signed by both parties (standard characteristic for contracts and documents). This might have limited customers who are treated as SLA customers, but lacked the correct documentation, out of scope. As well, it is by no means defined that only SLA customers in the case company get the attention regarding close monitoring of delivery reliability levels. Majority of the non-SLA customers are similarly important to case company than SLA-customers and possibly even treated with the same attention from the management side.

Validity of the research reflects the appropriateness of measures selected, accuracy of the research results and importantly, the generalization of the findings (Saunders et al. 2016, 202). In this research context it is believed that the validity of the measures used is extensively good. The dataset's shows actual delivery reliability data, and this is also one of the key components in this study. It can be said that measures set support the research well. For the accuracy perspective as well, the quantitative data and the data collection method leaves only little if any room for other interpretations. Results of the statistical analysis can be viewed as accurate and describing the research questions well.

Generalizability of the results is then more complex aspect to evaluate. In general, case studies are criticized as the results are not as generalizable. The dataset analyzed in this research that consisted of actual delivery reliabilities is quite extensive with 1275 observations. In terms of that large data, it can be said that data has many generalizable aspects and a lot of variation for example in customer size. Then again, this research as case study is solely focused in one organization setting and the results might not be generalizable for example for other industry settings, or for other companies. As Voss et al. (2002, 195) in their article mention, one of the challenges of case studies is having generalizable findings. To conclude the generalizability aspect, it can be said that yes, the results of this study have the potential to be generalized into other context as well, but further similar research is needed to support the results of this case study.

4 EMPIRICAL FINDINGS

This chapter presents the results of the statistical analyses that were justified and explained in detail in previous chapter. To help illustrate the findings, the critical to discussion results are displayed with supporting tables and figures, such as bar chart and histogram. First sub-chapter presents the descriptive statistics in the form of tables and explanations of those. Next sub-chapter focuses on the results of the t-tests, that are carried out to find out if the means of the groups are equal. In this research there were total of three t-tests conducted, for hypotheses H1, H2a and H2b. Last chapter displays the result from the Spearman's correlation analysis, for H3 hypothesis.

4.1 Descriptive tables

Descriptive tables are created with the intention to describe the data set as accurately as possible. Values of this table are number of observations, mean, standard deviation and minimum and maximum values. Variables from data to the descriptive table are presented next, with a short description of the definition of each:

- **Order lines total**, tells how many orders were delivered to a customer within review horizon of one month (count).
- **Missed deliveries total**, that tells the total number of how many of above-mentioned orders were not fulfilled in time (count).
- **Supplier hits**, which tells how many of the above-mentioned delays, were due to supplier not delivering the goods on time (count).
- **Supplier delivery reliability** tells the delivery reliability of supplier delivering the goods to warehouse on time (percent).
- **Warehouse hits**, indicates how many non-timely deliveries were due to warehouse delays (count).
- **Warehouse delivery reliability**, that indicates the performance of warehouse and what was the delivery reliability of that (percent).
- **Delivery reliability**, tells the overall delivery reliability, taken into consideration supplier hits and warehouse hits, so non-timely deliveries caused by supplier delays and warehouse delays. This figure is at the center of this research (percent).

In this chapter, there is two descriptive tables presented, first one with results of all variables and second table with the variables grouped based on the SLA and non-SLA group division. Below table is the firstly mentioned with all variables and measures in one table.

Table 7. Descriptive table for all observations.

ALL Variable	Obs	Mean	Std.Dev.	Min	Max
Order lines total	1,257	158.7749	194.0642	1	1195
Missed deliveries total	1,148	31.57404	48.40362	1	485
Supplier hits	1,139	29.59526	44.33935	1	485
Supplier delivery reliability	1,257	.7983801	.1895667	0	1
Warehouse hits	229	11.08297	26.70316	1	286
Warehouse delivery reliability	1,257	.9889491	.0465372	.4864865	1
Delivery reliability	1,257	.7873292	.1967477	0	1

There were a total of 1257 observations in the data set. Earlier it is mentioned that the data consisted of total 60 customers with an observation horizon of 24 months (2019-2020). If for all customer there would have been observable count for each month, that would have equaled total 1440 observations. Thus, on average, in this dataset there were around 21 observations per customer. When looking at the min and max of variable order lines total, it can be said that there was huge variation in number of order lines delivered to customers per month. As a minimum, there were only 1 order delivered per month and as a maximum there were as many as 1195 orders. Mean of the order lines total was 158 deliveries per observation month.

On average there were 31 missed deliveries per month. When comparing to the mean of order lines total, this is almost fifth of it, as if one in five deliveries was delayed, on average. The minimum value for missed deliveries is 1 and maximum value is 485. Number of observations of missed deliveries, when compared to order lines total tells, that there were only 109 observations where there were no delays at all.

Delivery reliability is scattered into three measures in this research context. In terms of supplier delivery reliability, the mean is 79,83 %. In terms of warehouse delivery reliability, the percent is remarkably higher, 98,89 %. Overall delivery reliability during the observation period of 2019-2020 on average is 78,73 %. As conclusion, it can be stated that warehouse delivery reliability is on the highest level. When looking at the supplier hits, there were on average 29 missed on-time deliveries, where min value being 1 and max value being 485, so same as for missed deliveries total. With regards to warehouse hits i.e., delays that occurred during the warehouse processes, the average was 11 orders missed per observed month. The weighing between supplier hits and warehouse hits can be observed by looking at the number of observations. As visible in Figure 13, for supplier delays, there were 1148 observation months, were the supplier delivery reliability performance fell back from the full 100 %. During the same time, warehouse delay observations totaled only at 229. Out of 1257 observation moths, there were delays in 1148 months. Which means that only for 109 out of 1257 customer observation months overall delivery reliability was full 100 percent.

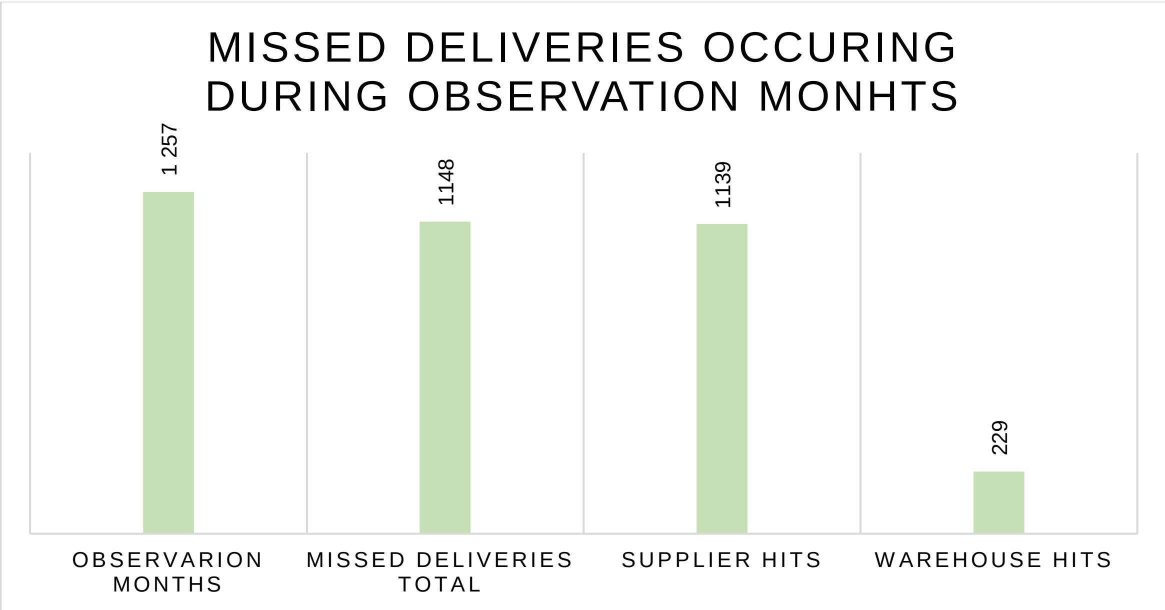


Figure 13. Missed on-time deliveries during observation months.

Next table presents the descriptive across the two comparison groups, which enables to have a closer look on the delivery performance differences across groups. Group 1 descriptive figures presented first and Group 2 descriptive figures second.

Table 8. Descriptive table by groups.

GROUP 1					
Variable	Obs	Mean	Std.Dev.	Min	Max
Order lines total	537	136.4655	226.0944	1	1151
Missed deliveries total	458	31.89301	53.10453	1	418
Supplier hits	456	30.33333	48.3912	1	349
Supplier delivery reliability	537	.76022821	.2264867	0	1
Warehouse hits	73	10.61644	35.29189	1	286
Warehouse delivery reliability	537	.9930797	.0339391	.6016713	1
Delivery reliability	537	.7533079	.2287214	0	1

GROUP2					
Variable	Obs	Mean	Std.Dev.	Min	Max
Order lines total	720	175.4139	164.4137	4	1195
Missed deliveries total	690	31.36232	45.05213	1	485
Supplier hits	638	29.10249	41.44372	1	485
Supplier delivery reliability	720	.8268351	.1504544	0	1
Warehouse hits	156	11.30128	21.68348	1	163
Warehouse delivery reliability	720	.9858683	.0538728	.4864865	1
Delivery reliability	720	.8127034	.1646599	0	1

As it can be observed, Group 1, where customers have SLA determining operations, is having a delivery reliability percent of 75,33 % on average. Group 2, which is the customer group without the SLA, has a delivery reliability of 81,27 % on average. When scattered into supplier and warehouse delivery reliabilities, Group 1 has a supplier delivery reliability of 76,02 % and warehouse delivery reliability of 99,30 %. On comparison Group 2 had a supplier delivery reliability 82,68 % and warehouse delivery reliability of 98,58 %. To conclude, overall delivery reliability and supplier delivery reliability is on better level for Group 2, but warehouse delivery reliability is slightly better for Group 1. All three delivery reliability measures for all, Group 1 and Group 2 are gathered for below Figure 14. From there it can be observed that the second measure, which is the warehouse delivery reliability is the highest figure for all three. As conclusion it can be said that it is the supplier delivery reliability that is lowering the overall delivery reliability measure and warehouse delivery reliability is enhancing it.

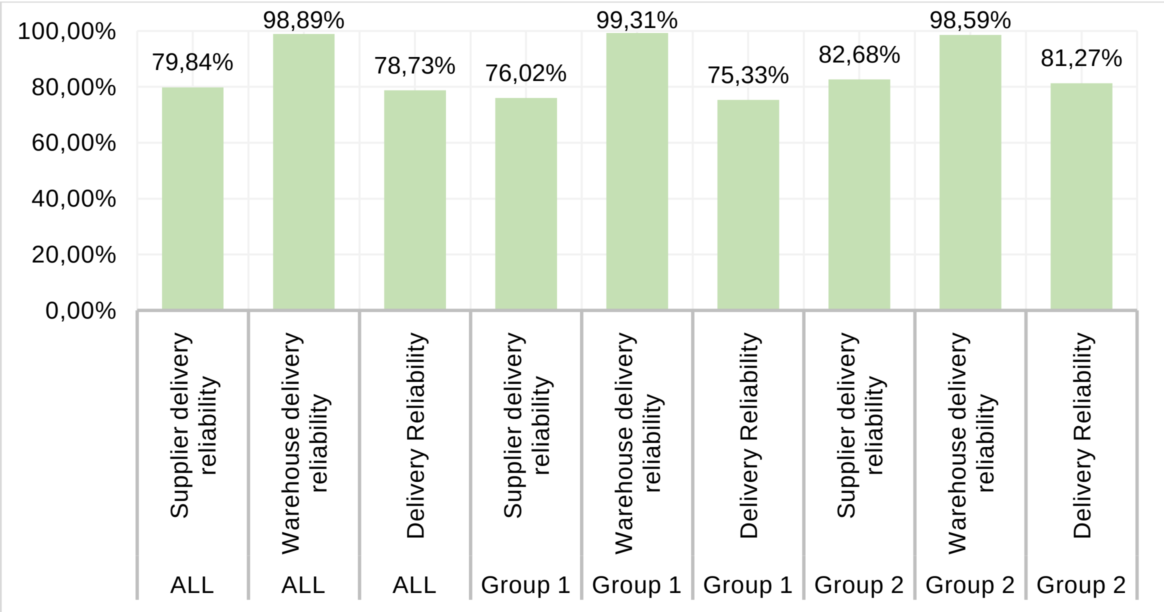


Figure 14. Delivery reliability figures, for all and based on grouping.

4.2 Testing for equal means across groups

4.2.1 Delivery reliability across groups

Before conducting the actual mean comparison by t-test, the equality of variances across groups needs to be tested, this part one phase is presented in below Table 9. Variable in question is delivery reliability, and the variance between the groups Group 1 and Group 2. Hypothesis setting is presented below the table, where H0 is accepted if p > 0.05.

Table 9. Part one: Test of equality of variances for variable delivery reliability.

Summary of Delivery reliability			
Group	Mean	Std.Dev.	Freq.
Group 1	.75330789	.22872136	537
Group 2	.81270337	.16465995	720
Total	.78732917	.19674766	1,257
W0 =	50.318324	df(1, 1255)	**Pr > F = 0.00000000**
W50 =	38.835299	df(1, 1255)	Pr > F = 0.00000000
W10 =	38.747420	df(1, 1255)	Pr > F = 0.00000000

H0: Variances are equal

HA: Variances are not equal

P-value in the equality of variances test is **0.00000000** which means that p-value is smaller than the required > 0,05 for discarding the null hypothesis. Therefore, H0 is rejected, and HA accepted. Variances across comparison groups for variable delivery reliability are not equal. This needs to be considered in the second part of t-test, by setting the variables as unequal variables in the Stata software. Part one being concluded with above statement; the test can be continued with the actual mean comparison test. The t-test's part two is where the means of the two groups are compared with statistical computing, for variable delivery reliability. Here again, the hypothesis is set as presented below, and H0 is accepted if p > 0.05.

Table 10. Part two: T-test of equal means for variable delivery reliability across groups.

Group	Obs	Mean	Std.Err.	Std. Dev.	[95% Conf. Interval]	
Group 1	537	.75330789	.0098701	.2287214	.7339192	.7726966
Group 2	720	.81270337	.0061365	.1646599	.8006557	.824751
combined	1,257	.7873292	.0055493	.1967477	.7764422	.7982162
diff		-.0593955	.0116222		-.0822043	-.0365867
diff = mean(Group 1) - mean(Group 2)			t = -5.1105			
Ho: diff = 0			Satterthwaite's degrees of freedom = 927.19			
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0000		**Pr(|T| > |t|) = 0.0000**		Pr(T > t) = 1.0000		

H0: Delivery reliability mean is the same for the sample groups

H1: Delivery reliability mean is not the same for the sample groups

This t-test with a p-value of **0.0000** discards the H0 hypothesis as p-value is smaller than 0,05 and concludes H1 to be accepted. There is statistical difference in the means of the two groups. By looking at the mean column in above table, it clearly indicates that the mean of Group 1 in terms of delivery reliability is lower .75330789 = 75,33 %, than for the comparison group, where the mean of delivery reliability is .81270337 = 81,27 %. The result is supported by the firstly presented descriptive tables.

4.2.2 Supplier delivery reliability across groups

Second t-test begins with testing for equality of variances for variable supplier delivery reliability. This variable indicates how well supplier has delivered the goods to case company warehouse before delivery to customer took place. Hypothesis setting can be seen after Table 11. Null hypothesis is to be accepted if p > 0,05.

Table 11. Part one: Test of equality of variances for supplier delivery reliability.

	Summary of Supplier delivery reliability		
Group	Mean	Std.Dev.	Freq.
Group 1	.76022821	.22648668	537
Group 2	.82683506	.15045437	720
Total	.79838011	.18956672	1,257
W0 =	64.515352	df(1, 1255)	**Pr > F = 0.00000000**
W50 =	47.388054	df(1, 1255)	Pr > F = 0.00000000
W10 =	47.102564	df(1, 1255)	Pr > F = 0.00000000

H0: Variances are equal

HA: Variances are not equal

P-value being **0.00000000** the null hypothesis can be rejected, and HA accepted. To conclude, the variances of variable supplier delivery reliability are not equal. This information is utilized in the second part of t-test continued with Stata software. The result of the H1 t-test have already been received, which sets some pre-assumptions to our succeeding t-tests. H1 resulted in a way that there is difference in the means across groups for variable delivery reliability. In this next t-test, the aim is to investigate further and scatter the complete delivery reliability into two pieces, supplier delivery reliability and warehouse delivery reliability. First one indicates the success rate of supplier delivering the goods to case company warehouse on time before the customer delivery takes place and latter one indicating if case company warehouse sent the goods on time. Hypothesis setting for the first mentioned is presented after Table 12.

Table 12. Part two: T-test of equal means of variable supplier delivery reliability across groups.

Group	Obs	Mean	Std.Err.	Std. Dev.	[95% Conf. Interval]	
Group 1	537	.7602282	.0097736	.2264867	.7410289	.7794275
Group 2	720	.8268351	.0056071	.1504544	.8158268	.8378433
combined	1,257	.7983801	.0053468	.1895667	.7878905	.8088698
diff		-.0666068	.0112678		-.0887219	-.0444918
diff = mean(Group 1) - mean(Group 2)			t = -5.9113			
Ho: diff = 0			Satterthwaite's degrees of freedom = 876.137			
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0000		**Pr(|T| > |t|) = 0.0000**		Pr(T > t) = 1.0000		

H0: Supplier delivery reliability mean is the same for the sample groups

H2a: Supplier delivery reliability mean is not the same for the sample groups

As it can be seen from above table, p-value being **0.0000** means again that the null hypothesis can be rejected and alternative hypothesis H2a applied. To conclude this finding, it can be stated that supplier delivery reliability mean is statistically not the same across the two groups. Supplier delivery reliability seems to be lower for Group 1 (mean .7602282 = 76,02 %) than for Group 2 (mean .8268351 = 82,68 %). This finding supports the result of the first t-test where the comparison was conducted for the overall delivery reliability.

4.2.3 Warehouse delivery reliability across groups

For the third and last t-test the hypothesis for part one test for equality of variances of variable warehouse delivery reliability is formulated as presented below Table 13. Null hypothesis expects the variances to be equal and alternative hypothesis that the variances are unequal. Hypothesis setting is seen after below table. H0 is rejected if p > 0,05.

Table 13. Part one: Test of equality of variance for warehouse delivery reliability.

	Summary of Warehouse delivery reliability		
Group	Mean	Std.Dev.	Freq.
Group 1	.99307967	.03393907	537
Group 2	.98586831	.05387277	720
Total	.98894906	.04653717	1,257
W0 =	20.7252566	df(1, 1255)	**Pr > F = 0.00000582**
W50 =	7.4237143	df(1, 1255)	Pr > F = 0.00652639
W10 =	9.1537637	df(1, 1255)	Pr > F = 0.00253257

H0: Variances are equal

HA: Variances are not equal

Result is a p-value of **0.00000582** which means again, that H0 is rejected, and HA is accepted. Variances of variable warehouse delivery reliability are not equal. With this information it is proceeded to the second part of the t-test. Part two is a test of equal means for variable warehouse delivery reliability. Hypothesis is set as before; null hypothesis expects that the means are same and alternative hypothesis that means are not the same in the two groups. The value of null hypothesis rejection is p > 0,05.

Table 14. Part two: T-test of equal means of variable warehouse delivery reliability across groups.

Group	Obs	Mean	Std.Err.	Std. Dev.	[95% Conf. Interval]	
Group 1	537	.9930797	.0014646	.0339391	.9902027	.9959567
Group 2	720	.9858683	.0020077	.0538728	.9819266	.98981
combined	1,257	.9889491	.0013126	.0465372	.9863739	.9915242
diff		.0072114	.0024851		.0023357	.012087

diff = mean(Group 1) - mean(Group 2)	t = 2.9018	
Ho: diff = 0	Satterthwaite's degrees of freedom = 1223.18	
Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.9981	**Pr(|T| > |t|) = 0.0038**	Pr(T > t) = 0.0019

H0: Warehouse delivery reliability mean is the same for the sample groups

H2b: Warehouse delivery reliability mean is not the same for the sample groups

P-value for this final t-test is **0.0038** and this supports in rejecting the null hypothesis and accepting the H2b hypothesis. According to this statistical result, it can be said that warehouse delivery reliability is not statistically same for the two comparison groups. Unlike in two preceding tests, Group 1 was identified to have lower mean in terms of delivery reliability and supplier delivery reliability. Here on the contrary, Group 1 mean in terms of warehouse delivery reliability is slightly higher than for Group 2. This indicates that customer group where the SLA is in use, have better warehouse delivery reliability than the customer group without SLA determining operations.

4.3 Testing for correlation between variables

A spearman's correlation analysis was conducted to find out if variables order lines total and missed deliveries total are related to one another. In other words, does high number of orders per month, indicate higher number of missed deliveries. Spearman's correlation analysis begins with test for normal distribution of the data. Based on the first p-value result, correlation analysis method is selected. As mentioned earlier, Spearman is adapted in case of non-normally distributed data. Probability, p-value is set for 0.05. Hypothesis for Shapiro-Wilk test for data distribution is seen after the Table 15.

Table 15. Shapiro-Wilk test for normal distribution.

	Shapiro-Wilk W test for normal data				
Variable	Obs	W	V	z	Prob>z
Order lines total	1,257	0.72631	212.618	13.395	**0.00000**
Missed deliveries total	1,148	0.59288	291.176	14.130	**0.00000**

H0: Variable is normally distributed

HA: Variable is not normally distributed

Probability, p-value **0.00000** being under the 0.05 for both variables order lines total and missed deliveries total shows, that the H0, null hypothesis can be rejected for both, and alternative hypothesis accepted. Neither of the variables are normally distributed. This may also be observed from the histograms presented in next page Figure 15 and Figure 16. Where in case of normal distribution, the curve would have been at bell shape, rather than what is now seen in the graph.

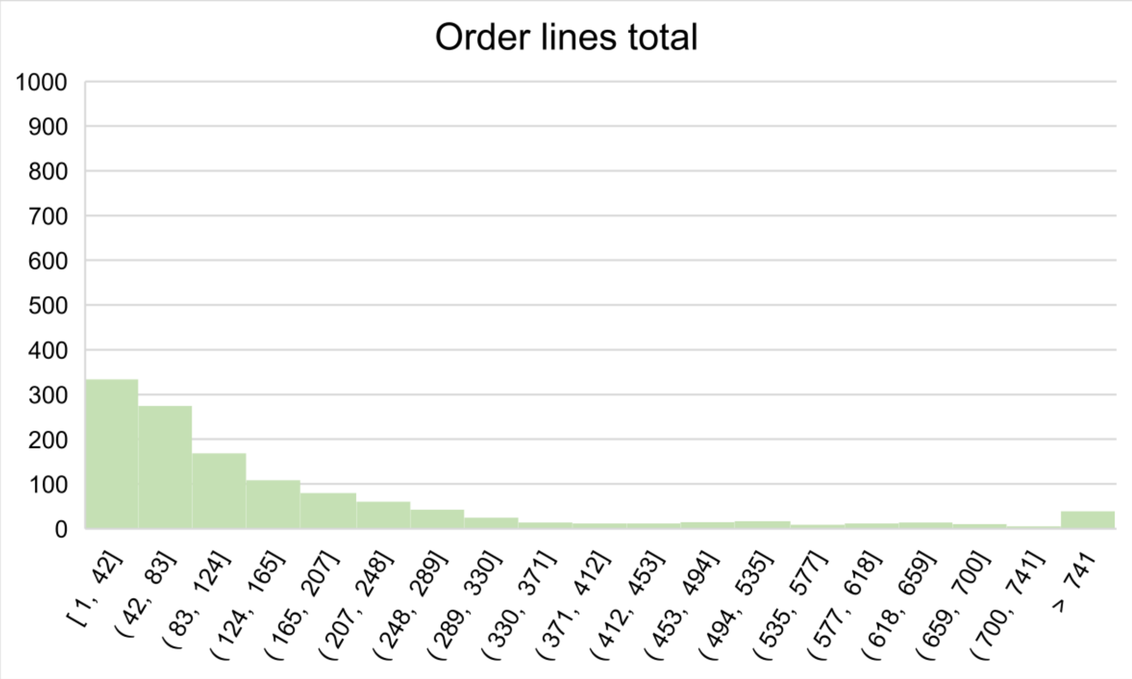


Figure 15. Non-normal distribution of variable order lines total.

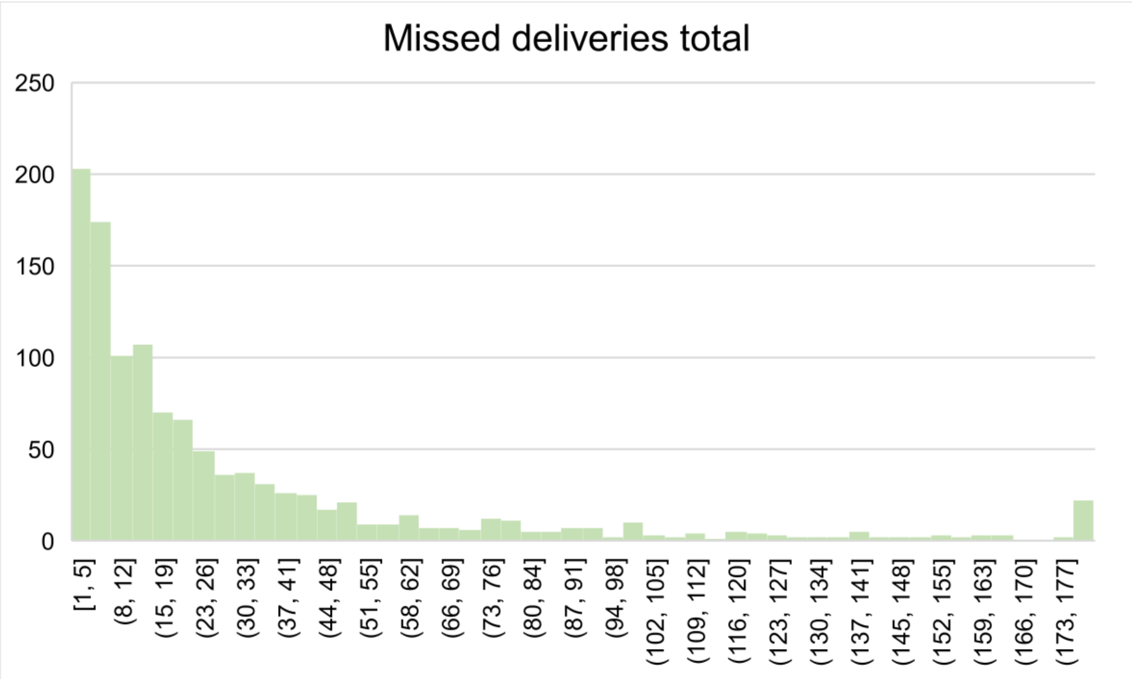


Figure 16. Non-normal distribution of variable missed deliveries total.

As we have now proved the variables to be non-normally distributed, a Spearman’s correlation analysis can be selected and carried out. Here the aim is to test the strength of relationship of variables order lines total and missed deliveries total. Spearman’s correlation gives two results. First the null hypothesis is accepted if p > 0.05. Secondly, a Spearman's rho correlation is detected between absolute values of -1 and 1. Where -1 concerns the correlation to be perfect negative and 1 reflects perfect positive correlation. Value 0 means there is no correlation.

Table 16. Spearman’s correlation.

Number of obs =	1148
Spearman's rho =	**0.7445**
Test of Ho: Order lines total and Missed deliveries total are independent	
Prob > |t| =	**0.0000**

H0: There is no correlation between variables
H3: There is correlation between variables

Spearman’s correlation analysis shows, with a p-value of **0.0000** that the null hypothesis is rejected and hypothesis with the statement that there is correlation between the variables, accepted. Correlation coefficient shows a value of **0.7445**, which means that there is in fact strong positive correlation between the variables. To conclude the result of this analysis, it can be said that the higher the number of order lines, the higher the number of missed deliveries.

4.4 Summary of the results

To conclude empirical findings, conceptual framework was revised for Figure 17 with the results of the statistical analyses. From the figure it may be observed, that for all the hypotheses, null hypothesis with the assumption “no difference found” is rejected and alternative hypothesis “there is difference” accepted. When it comes to the t-tests H1, H2a and H2b the results are in line with each other. As the H1 was accepted and found out that the means of the groups are not the same when it comes to delivery reliability, it naturally gave some pre-assumptions for preceding t-tests. Later tested H2a and H2b support the findings of the H1, by the result determining that for the supplier delivery reliability and warehouse delivery reliability as well, the means are not the same across groups.

Spearman’s correlation test for H3 gives an insightful result; while the number of orders increase, so does the number of missed deliveries. Here as well the null hypothesis is rejected, and alternative hypothesis accepted. This result supports in finding perspectives regarding the reasons on the background of falling back from the perfect delivery reliability and while assessing the answers to research questions.

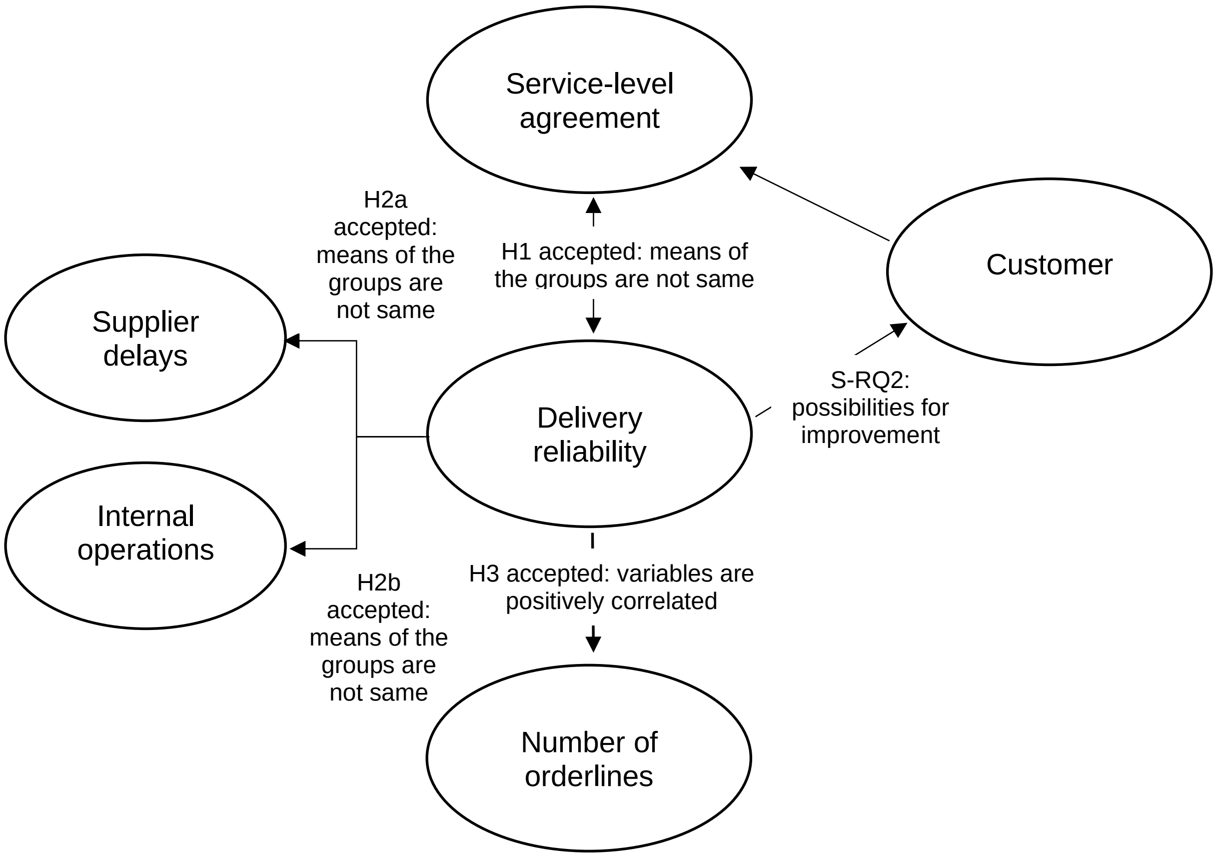


Figure 17. Conceptual framework revised with results.

5 DISCUSSION

This chapter discusses the findings of this research by providing answers to research questions set at the beginning of this thesis. Aim is to analyze each question at a time and provide responses in the light of the literature review, presented in Chapter 2 and empirical results from Chapter 4. As well, managerial implications and suggestion for further research are discussed.

5.1 Answers to research questions

In this sub-chapter research questions are presented one at a time with an answer in the light of the empirical results and theoretical background. Starting with the main research question and moving on to the two sub-research questions.

Main research question (RQ)

What is the impact of service-level agreement to delivery reliability?

Based on the empirical findings, it may be addressed that service-level agreement possibly affects the delivery reliability, but the impact is by no means positive. Result of the t-test for variable delivery reliability show, that there is statistical difference across group means for SLA customers and non-SLA customers when it comes to delivery reliability. When looking at the details of the t-test script and descriptive table presented at the empirical analysis chapter, it can be observed that the overall delivery reliability is lower for customers with SLA than for non-SLA customers. According to the statistical analyses over the review horizon, delivery reliability mean for SLA customers was 75,33 % whereas for the customers who do not have SLA, had a delivery reliability mean of 81,27 %. When looking at all the customers together, delivery reliability is 78,73 % on average, which is then slightly better than just for the SLA customers. The reasons behind this finding are discussed next. Possible motives for establishing service-level agreements are identified and classified under two views, which are,

- customer motives
- case company as supplier motives

What is the possible reason for delivery reliability not being better for SLA companies? To answer this question, a review to the literature around SLA's in in place. As discussed in the theory chapter, SLA is the commitment supplier takes when commencing business practices with the buying company (Katok et al., 2008, 609). As well, by SLA's buying company expects certain performance levels or other requirements from the supplier (Chen, 2018, 1338). Such as product availability, quality, and timely deliveries (Liang & Atkins, 2013, 1103). Enhancing supplier's performance is the main driver in initiating SLA's with suppliers (Chen & Thomas, 2018, 553). Maestrini et al. (2018, 2056) study show, that supplier performance measurement has an impact on the improvement of supplier performance.

Considering above, one possible answer is that buying company was motivated to establish SLA with supplier, due to already low performance in the past, and there has not been significant improvement after the SLA establishment. This suggestion also relates to the supplier development practices, where general view is that if supplier is performing on non-accepted level, buying company has two ways to improve supplier performance, either by switching the supplier or by development activities (Krause & Ellram, 1997, 21-22). Setting-up SLA may have been one of the improvement activities. Alternative option is that SLA requirement from the customer may have been just a standard way when engaging into operations with the supplier.

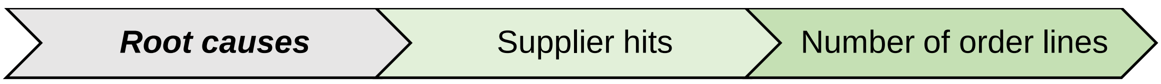
SLA may not always be the buying company requirement, as well, case company as supplier steers agreements with is customers, and above presented motivations apply to these situations as well. As mentioned while discussing supplier performance measurement systems in the literature review, communication between parties and transferring the data to supplier regarding performance are essential components of the supplier performance measurement (Maestrini et al., 2018, 304). In case of poor performance is detected from the evaluation, supplier takes an interactive approach to performance review received (Maestrini et al., 2018, 304). This may contribute as initiating SLA and steering towards alignment agreements with the customer. Or supplier is motivated to perform on a sufficient level simply to keep the business with the customer (Chen & Thomas, 2018, 553). Supplier might as well have marketing and sales related incentives at interest when commencing into SLA agreements. Alternatively, supplier has a target of becoming a strategic supplier for their customer and thus improving the relationship and operations even further.

Customers in the SLA group had on average fewer order lines per observation month than the non-SLA customers. The result of the correlation test showed, that while the number of order lines increased, so did the number of missed deliveries. Considering this, it would be logical to draw the conclusion that the bigger the customer number of deliveries wise, the worse the delivery reliability performance would be. But this assumption does not apply when evaluating why delivery reliability is not as good for the SLA customers than for the non-SLA customers.

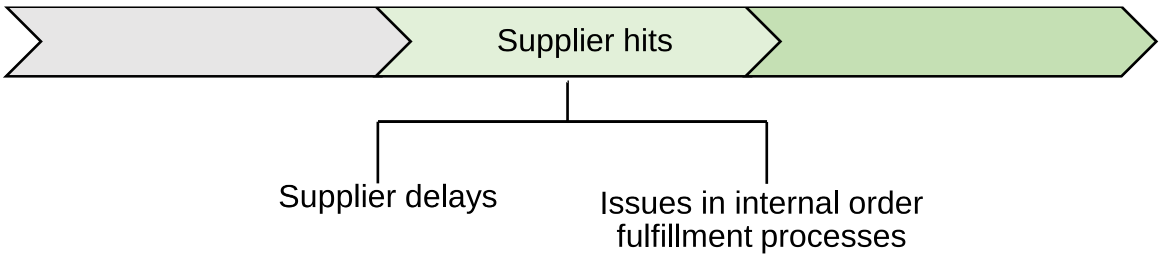
In case of possible penalties or other consequences (Chen, 2018, 1329, 1337) determined in the SLA, supplier would avoid falling below the performance expectations at all costs. If the situation were depending on this fact, it would logically thinking imply better delivery performance for the SLA companies. But as Ching et al., (2011, 5176-5177) note, in cases of penalties in place, fixed penalties diminish supplier motivation for good delivery performance due to the fee usually being relatively low. Therefore, this theory as well may lose its arguments and it is worth concluding that possible consequences are not the critical factor that would imply better delivery reliability for the SLA customers.

Sub-research question 1

What is the cause of falling below delivery reliability expectations?



Results of the empirical analysis show, that there are two main reasons why performance in terms of delivery reliability is falling below full 100 percent. First reason are delays caused by goods not being available on time on stock before the delivery to customer took place. Here in this research context referred to as supplier hits. Second reason is the number of orders to customers during the observation month. It seems, according to the results, that the more orders to customer were each month, the more missed on-time deliveries were too. Next possible reasons are analyzed in detail starting with supplier hits and moving on to the number of order lines.



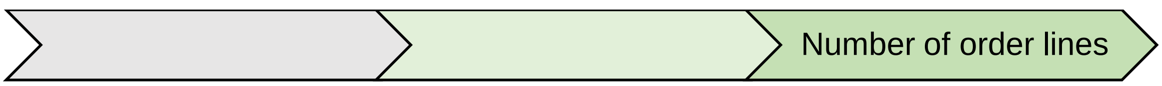
As the result of t-test for shows, there is statistical difference between the means of SLA group and non-SLA group in terms of supplier delivery reliability measures. When looking at the details of the t-test script and descriptive table, supplier delivery reliability mean is 82,68 % for the better performance having group of non-SLA customers. For the SLA customer group corresponding supplier delivery reliability mean is 76,02 %. Descriptive table shows that there were as many as 1139 observation months out of the total 1257 where there was at least one missed delivery due to supplier delay. There may be diverse reasons behind supplier delivery reliability falling below the full 100 percent. Supplier hits in this research were in general considered as missed deliveries due to goods not being on stock on time before the delivery to customer took place. Therefore, all other reasons in addition to actual supplier delays fall under this same supplier delivery reliability measure. For example, issues at internal processes such as order acceptance and allocation planning processes might as well impact goods availability at the time the delivery to customer took place.

Supplier delays are causing most delays in this research context and this is the main driver that depresses delivery reliability. Good performance of the supplier is connected to the success of entire supply chain (Kannan, 2018, 392) and it is the aim of correct supplier choice to ensure quality standards being fulfilled, optimal cost and timely deliveries to the purchasing company (Yang et al., 2017, 105). One main scenario why supplier delays are the main cause of lowering delivery reliability, is that suppliers of the case company are not performing on sufficient level.

As case company mostly operates make-to-order basis, supplier delivering goods not on time has direct impact on the delivery reliability of case company towards its customers. Supplier delivery reliability is measured by supplier measurement systems, that evaluates efficiency and effectiveness of supplier’s actions (Maestrini et al., 2018, 298). Main purpose of supplier evaluation is to first identify correct metrics to evaluate suppliers with and then to measure and monitor supplier performance. (Park et al., 2010, 501-502). It is possible that case company is only starting supplier performance measurement activities, is not having correct attention on that, or is struggling with suppliers of bottleneck item categories, where markets are supplier dominant (Park et. al (2010, 500).

Case company as supplier is forced to consider demand risks that occur mostly because of fluctuation in customer demand. Risk mitigation perspective is to not end up with excessive inventory or have stock-outs to customer. (Diabat et al., 2012, 3039). Therefore, the balance of mitigating demand risks and executing order fulfillment to the best extent may result in one of these not being on the perfect level. With the perfect inventory level and incoming flow, all open orders can be delivered from the available supply and if not, allocation planning as integral order fulfillment process comes in question (Seitz, et al.,2020, 1). Allocation planning is conducted by considering customer service-level agreements (Thomas, 2005, 80), but for this research context this seems not to be the case, or at least allocation is not increasing the level of delivery reliability for the service-level agreement determined customers.

Another operational order fulfillment process that is potential root-cause for the delays are possible issues in the process of order confirmation (Kritchanchai & MacCarthy, 1999, 818). For company operating make-to-order (MTO) basis, it is characteristics that orders from the customers are defined by production lead times (Ogˇuz et al., 2010, 200-201). When discussing MTO company context as case-company is, issues at ordering according to production lead times cannot be discarded from root-causes. These two aspects of order fulfilment processes are seen as potentially causing delays that are in this research defined under the supplier delivery performance measure. In addition to previously presented actual supplier delays and actions to mitigate that, the root-causes may lie in the case company operational processes.



In addition to supplier hits, number of orders is seen to increase the number of missed deliveries, and through that, effect on delivery reliability. Result of the correlation test show, that there is strong positive correlation for variables order lines total and missed deliveries total. Which means that as the number of orders per observation month increased, so did the number of missed deliveries. As seen from the descriptive table, the average number of order lines per month was 158 orders. Minimum value 1 order per month and a maximum value of 1195 orders per month. The result of the correlation is logically explainable. For example, in a case of as many as the maximum value, 1195 order lines, the likelihood of some of the order lines being late is more valid than if the customer had only the minimum value of 1 order line to be delivered. Vast number of orders per month indicates comprehensive business relation and different goods from different categories being delivered to customer. This naturally implies many more suppliers and supply chain partners involved in the on-time delivery and more suppliers and their performance to evaluate and manage.

Sub-research question 2

How is the case company performing as supplier in terms of delivery reliability?

Case company's performance as supplier towards its customers is on a sufficient level and close to customer expectations in terms of delivery reliability. Overall delivery reliability is 78,73 % which consists of both supplier delivery reliability and warehouse delivery reliability measures. Especially it is worth mentioning, that warehouse delivery reliability, which is considered as internal operational process, is exceptionally good and close to full 100 percent. This finding is analyzed further in next chapters. For the customers in this research, warehouse delivery reliability mean was 98,89 % where warehouse hits occurred only for 229 months out of 1257 observation months. It was mentioned while discussing this research data collection that customers in the SLA group had delivery reliability expectations between 80 and 100 percent. If this was considered as standard customer expectation and reflected to the delivery reliability findings of this research, it can be said that delivery reliability is close to expectations, but still at the bottom of the scale of customer expectations.

While supplier delays and number of orders seemed to be the main drivers in this study context to affect delivery reliability, warehouse delays seemed to have very little impact on falling delivery performance. T-test for variable warehouse delivery reliability shows that there is statistical difference between the two comparison groups mean for warehouse delivery reliability. Warehouse delivery reliability was observed to be slightly better for the SLA group customers, according to the descriptive table figures and the t-test script for variable warehouse delivery reliability. Although the difference in mean, when compared to the non-SLA group was very minimal. For both groups warehouse delivery reliability was close to full 100 percent. For the SLA group warehouse delivery reliability was 99,30 % and for the comparison group of non-SLA customers 98,58 %.

As mentioned in the literature review chapter when discussing risks impacting delivery reliability, according to Tang & Tomlin (2008, 13) company's inbound and outbound logistics operations are a source of risks and may impact company's effectiveness. Empirical results indicate that warehouse delivery reliability is on a very good level, close to perfect 100 percent, and delays due to warehouse being late with their processes occur only randomly. This may indicate that operational risks are well managed and mitigated at the case company organization. As well, delays in warehouse inbound and outbound processes have direct impact in customer delivery delays, which seems to be mitigated very well in the case company. This statement is supported by the number of observations in the descriptive table for variable warehouse hits. Where only 229 observation months out of total 1257 months were impacted by warehouse delays.

There are improvement possibilities in terms of increasing delivery reliability level. Improvable measure that needs further attention is supplier delivery reliability and processes behind that, these recommendations were discussed more in depth while answering sub-research question 1. Supplier delivery reliability is much more than just measuring on-time deliveries. As defined in the supplier chain operations reference (SCOR) model, as well delivery of correct product, at the correct time, in the correct form in terms of packaging and condition, with the correct documentation provided and to the correct customer are considered as components of delivery reliability (Cirtita & Glaser-Segura, 2012, 301). For the case company this means that although the overall delivery reliability in terms of on-time deliveries, was identified to have some improvement possibilities, it may be that other components of delivery reliability are well achieved.

It is highly possible that case company customers whose delivery reliability is not on accepted level, have noticed this and initiated supplier development actions, that aim at improving the supplier performance (Park et al., 2010, 506). As mentioned earlier, case company as supplier gives great emphasis to following SLA requirements. Any below expectations performance in terms of delivery reliability for customers who measure, report, and share the evaluation with the supplier, is observed and noted in the case company and as well in at customer's side.

5.2 Conclusions

Key finding in this thesis is that service level agreements have an impact on the delivery reliability, but not on the positive direction. There are possible ulterior motives for companies, both supplying company and purchasing company to establish SLA agreement for the business operations. This thesis gives a good overall understanding of the nature of SLA's and impact of those to business operations and a good direction from the empirical point of view, where further studying may begin from. The actual motives are not in focus in thesis context, due to for this research context selected dataset not supporting in finding these answers.

This study shows that there is an actual need for investigating the relation of SLA's and delivery reliability more. Pre-assumption at the beginning of this thesis writing process and even after the literature review, was that SLA customers would somehow naturally have better delivery reliability than customers who have no SLA requirements. This was proven false in the light of the empirical findings. This discussion chapter where the answers to research questions were presented as well with the conclusion, focused solely on analyzing the possible causes for the contrary to pre-assumption findings. Results were to some extent surprising but while analyzing the possible causes, theory around the concepts was emerged in different light and perspective and founded supporting the empirical findings. Therefore, it may be concluded that literature review and empirical findings are in align and that theory supports later analyzing. There is potential for the generalizability of the results, but there is strong need for further investigation and recreation the research for different context and industry.

Performance of the supplier is close to sufficient level, but some improvements aspects are identified. Research results show that overall delivery reliability being 78,73 % is slightly below and on the bottom of the level of customer expectation scale of 80 to 100 percent. Supplier hits is the biggest cause for falling below the performance level expectations. Case company internal operations success is measured by warehouse delivery reliability, which is on an excellent level, being close to full 100 percent. In addition to delays caused by supplier, number of order lines per observation month was identified to be the cause of delayed on-time deliveries.

5.3 Managerial recommendations

Key take-away for managerial recommendation is to investigate the processes that impact supplier delivery reliability. This research was not able to scatter supplier delivery reliability measure into sub-processes but identified the need to have a closer attention on that. Alternative reason for delays in addition to supplier delays are internal processes. As well, it is important to identify actual supplier delays that cause negative effect on delivery reliability for case company's customers. By this it is possible to design, implement and review supplier performance measurement systems and to determine the supplier relationship management actions that aim at improving supplier performance.

Another recommendation is to develop service-level agreement use and clearly define the required outcome of those agreements. Service-level agreement components include review horizon, expected service-level in terms of on time deliveries and other measurements and the consequence, financial or non-financial, of not being able to respond to determined expectations. All components should be utilized in the best possible way and set to support the supplier-buyer relationship. As well, to improve and streamline the operations of both parties. There is a growing need to make alignments and agreements with customers regarding expectations. Such agreements are considered to be at the interest of both supplier and buyer and to drive mutual benefits.

5.4 Limitations and suggestion for future research

This study was conducted in one company context and to be more specific, one division and geographical limitation of the company business operations. Thus, the results are especially relevant for this specific division while evaluating overall performance. Results cannot be guaranteed to apply fully to other company or even case company's other division contexts. The nature of the business the case company operates in is to some extent unique, and possibly does not have similar characteristics and challenges than other industries.

There are other requirements purchasing company expects from supplier in addition to certain delivery reliability level (Chen, 2018, 1338). In this research these were not studied, due to the difficult to quantify aspect. There was no available data regarding other requirements and as it was chosen to focus solely on the delivery reliability measure. Then again delivery reliability consists of much more than just on-time deliveries (Cirtita & Glaser-Segura, 2012, 301), and other aspects were not studied due to same reasons. Further research could address these points. It would be insightful to study other customer expectations and delivery reliability components. There might be difficulties in quantifying the results, thus qualitative methods would support this kind of research set-up.

This research was viewed from the supplying company perspective. First recommendation for further research would be to conduct similar research from the purchasing company perspective. Benefit of this kind of research would be exact knowledge regarding delivery reliability expectations and this could be reflected and analyzed for all supplier performances. Case company of this research is a supplier that operates both make-to-stock and make-to-order basis. Alternatively, it would be interesting to do further research in make-to-stock manufacturer context. As this would exclude supplier delays, there would possibly be other root causes for delivery reliability was falling below full 100 percent. In this type of research focus would be switched to inventory planning, excess stock avoidance and allocation rules in case of scarce supply. Lastly, there is robust need to conduct similar research for other case-study contexts, possibly in the same industry or other industries where SLAs are in use. Even for other divisions in the case company to have cross reference results and combine the findings to the benefit of many.

LIST OF REFERENCES

Aljandali, A. (2016) Quantitative Analysis and IBM® SPSS® Statistics: A Guide for Business and Finance. [Book] Cham: Springer International Publishing AG.

Altendorfer, K. & Jodlbauer, H. (2011) An analytical model for service level and tardiness in a single machine MTO production system. International journal of production research. [Online] 49 (7), 1827–1850.

Amer, Y. et al. (2010) Case study: Optimizing order fulfillment in a global retail supply chain. International journal of production economics. [Online] 127 (2), 278–291.

Arroyo-López, P., Holmen, E. and de Boer, L. (2012) How do supplier development programs affect suppliers?: Insights for suppliers, buyers and governments from an empirical study in Mexico. Business Process Management Journal. [Online] 18(4) 680– 707.

Baghalian, A. et al. (2013) Robust supply chain network design with service level against disruptions and demand uncertainties: A real-life case. European journal of operational research. [Online] 227 (1), 199–215.

Baskaran, V., et al. (2012) Int . J . Production Economics Indian textile supplier's sustainability evaluation using the grey approach. Intern. Journal of Production Economics. Elsevier. [Online] 135(2) 647–658.

Bertsimas, D. & Paschalidis, I. C. (2001) Probabilistic Service Level Guarantees in Make-to-Stock Manufacturing Systems. Operations research. [Online] 49 (1), 119–133.

Bourne, M. and Mills, J. (2000) Designing, implementing, and updating performance measurement systems, 20(7), [Online] 754–771.

Brabazon, P. G. & MacCarthy, B. L. (2017) The automotive Order-to-Delivery process: How should it be configured for different markets? European journal of operational research. [Online] 263 (1), 142–157.

Cees, J., Weele, V. and Arjan, J. (2005) Purchasing Portfolio Models : A Critique and Update. [Online]

Chen, C. (Jimmy) & Thomas, D. J. (2018) Inventory Allocation in the Presence of Service-Level Agreements. Production and operations management. [Online] 27 (3), 553–577.

Chen, C.-M. (Jimmy) (2018) A review and analysis of service level agreements and chargebacks in the retail industry. The international journal of logistics management. [Online] 29 (4), 1325–1345.

Ching, W.-K. et al. (2011) Inducing high service capacities in outsourcing via penalty and competition. International journal of production research. [Online] 49 (17), 5169–5182.

Chopra, S., M. S. Sodhi. 2004. Managing risk to avoid supply-chain breakdown. MIT Sloan Manage. [Online] Rev.46(1): 53–62

Christopher, M. & Lee, H. (2004) Mitigating supply chain risk through improved confidence. International journal of physical distribution & logistics management. [Online] 34 (5), 388–396.

Christou, I. T. & Ponis, S. (2009) A hierarchical system for effective coordination of available-to-promise logic mechanisms. International journal of production research. [Online] 47 (11), 3063–3078.

Cirtita, H. & Glaser-Segura, D. A. (2012) Measuring downstream supply chain performance. Journal of manufacturing technology management. [Online] 23 (3), 299–314.

Diabat, A. et al. (2012) Supply chain risk management and its mitigation in a food industry. International journal of production research. [Online] 50 (11), 3039–3050.

Durugbo, C. et al. (2014) Managing integrated information flow for delivery reliability. Industrial management + data systems. [Online] 114 (4), 628–651.

Friedl, G. and Wagner, S. M. (2012) Supplier development or supplier switching? International Journal of Production Research. [Online] 50(11) 3066–3079.

Geher, G. & Hall, S. (2014) Straightforward Statistics: Understanding the Tools of Research. Cary: Oxford University Press, Incorporated. [Online]

Gopal, P. R. C. and Thakkar, J. (2011) A review on supply chain performance measures and metrics. [Online] 2000-2011.

Gurvich, I. et al. (2008) Service-Level Differentiation in Call Centers with Fully Flexible Servers. Management science. [Online] 54 (2), 279–294.

Hadj Youssef, K. et al. (2004) Efficient Scheduling Rules in a Combined Make-to-Stock and Make-to-Order Manufacturing System. Annals of operations research. [Online] 126 (1), 103–134.

Henri, J. F. (2006) Management control systems and strategy: A resource-based perspective, Accounting, Organizations and Society, 31(6), pp. 529–558.

Hopp, W. J. et al. (1997) Easily Implementable Inventory Control Policies. Operations research. [Online] 45 (3), 327–340.

Infanger, D. & Schmidt-Trucksäss, A. (2019) P value functions: An underused method to present research results and to promote quantitative reasoning. Statistics in medicine. [Online] 38 (21), 4189–4197.

Jiang, Y. et al. (2019) Service Level Constrained Inventory Systems. Production and operations management. [Online] 28 (9), 2365–2389.

Jonsson, P. et al. (2007) Applying advanced planning systems for supply chain planning: three case studies. International journal of physical distribution & logistics management. [Online] 37 (10), 816–834.

Katok, E. et al. (2008) Inventory Service-Level Agreements as Coordination Mechanisms: The Effect of Review Periods. Manufacturing & service operations management. [Online] 10 (4), 609–624.

Kloos, K. & Pibernik, R. (2020) Allocation planning under service-level contracts. European journal of operational research. [Online] 280 (1), 203–218.

Krause, D. R. and Ellram, L. M. (1997) Critical elements of supplier development: The buying-firm perspective, European Journal of Purchasing and Supply Management. [Online] 3(1) 21–31.

Krause, D. R., Handfield, R. B. and Scannell, T. V. (1998) An empirical investigation of supplier development: Reactive and strategic processes. Journal of Operations Management. [Online] 17(1) 39–58.

Kritchanchai, D. & MacCarthy, B. (1999) Responsiveness of the order fulfilment process. International journal of operations & production management. [Online] 19 (8), 812–833.

Koufteros, X. et al. (2014) The effect of performance measurement systems on firm performance: A cross-sectional and a longitudinal study. Journal of operations management. [Online] 32 (6), 313–336.

Lambert, D. M. and Schwieterman, M. A. (2012) Invited paper Supplier relationship management as a macro business process. [Online] 337–352.

Lambert, D.M. (2008), Supply Chain Management: Processes, Partnerships, Performance, 3rd ed., Supply Chain Management Institute, Sarasota, FL.

Liang, L. & Atkins, D. (2013) Designing Service Level Agreements for Inventory Management. Production and operations management. [Online] 22 (5), 1103–1117.

Maestrini, V., et al. (2017a) Supply chain performance measurement systems: A systematic review and research agenda, Intern. Journal of Production Economics. Elsevier, 183(August 2015) [Online] 299–315.

Maestrini, V., et al. (2017b) The impact of supplier performance measurement systems on supplier performance: A dyadic lifecycle perspective, International Journal of Operations and Production Management. 38(11) [Online] 2040–2061.

Maestrini, V., et al. (2018) Supplier performance measurement systems : Communication and reaction modes, Industrial Marketing Management. Elsevier. [Online] 74(March 2017) 298–308.

Melnyk, S. A. et al. (2014) Is performance measurement and management fit for the future ?, Management Accounting Research. Elsevier Ltd, 25(2) [Online] 173–186.

Milner, J. M. & Olsen, T. L. (2008) Service-Level Agreements in Call Centers: Perils and Prescriptions. Management science. [Online] 54 (2), 238–252.

Min, S. et al. (2019) Defining Supply Chain Management: In the Past, Present, and Future. Journal of business logistics. [Online] 40 (1), 44–55.

Nooraie, S. V. & Mellat Parast, M. (2015) A multi-objective approach to supply chain risk management: Integrating visibility with supply and demand risk. International journal of production economics. [Online] 161192–200.

Ogˇuz, C. et al. (2010). Order acceptance and scheduling decisions in make-to-order systems. International journal of production economics. [Online] 125 (1), 200–211.

Okongwu, U. et al. (2012) A decision support system for optimising the order fulfilment process. Production planning & control. [Online] 23 (8), 581–598.

Park, Jongkyung et al. (2010) An integrative framework for supplier relationship management. Industrial Management and Data Systems. [Online] 110(4) 495–515.

Patrucco, A. S. et al. (2021) Does relationship control hinder relationship commitment? The role of supplier performance measurement systems in construction infrastructure projects. International journal of production economics. [Online] 233108000–.

Prahinski, C. & Benton, W. . (2004) Supplier evaluations: communication strategies to improve supplier performance. Journal of operations management. [Online] 22 (1), 39–62.

Rajagopalan, S. (2002) Make to Order or Make to Stock: Model and Application. Management science. [Online] 48 (2), 241–256.

Romule, K. et al. (2019) Supplier performance assessment. Benchmarking : an international journal. [Online] 27 (2), 817–838.

Sahin, F. & Robinson, E. P. (2005) Information sharing and coordination in make-to-order supply chains. Journal of operations management. [Online] 23 (6), 579–598.

Saunders, M. et al. (2016) Research methods for business students. Seventh edition. Harlow, Essex: [Book] Pearson Education.

Sawik, T. (2009) Multi-objective due-date setting in a make-to-order environment. International journal of production research. [Online] 47 (22), 6205–6231.

Seitz, A. et al. (2020) Data driven supply allocation to individual customers considering forecast bias. International journal of production economics. [Online] 227107683–.

Sieke, M. A. et al. (2012) Designing Service Level Contracts for Supply Chain Coordination. Production and operations management. [Online] 21 (4), 698–714.
Simpson, P, et al. (2002) Measuring the performance of suppliers : An analysis of evaluation processes. [Online]

Sodhi, M. S. (2005) Managing Demand Risk in Tactical Supply Chain Planning for a Global Consumer Electronics Company. Production and operations management. [Online] 14 (1), 69–79.

Sodhi, M. S. et al. (2012) Researchers' Perspectives on Supply Chain Risk Management. Production and operations management. [Online] 21 (1), 1–13.

Soepenberg, G. et al. (2012) A framework for diagnosing the delivery reliability performance of make-to-order companies. International journal of production research. [Online] 50 (19), 5491–5507.

Sridharan, V. & Li, X. (2008) Improving delivery reliability by a new due-date setting rule. European journal of operational research. [Online] 186 (3), 1201–1211.

Stadtler, H. (2005) Supply chain management and advanced planning—basics, overview and challenges. European Journal of Operational Research. [Online] 163 (3), 575–588.

Tang, C. & Tomlin, B. (2008) The power of flexibility for mitigating supply chain risks. International journal of production economics. [Online] 116 (1), 12–27.

Thomas, D. J. (2005) Measuring Item Fill-Rate Performance in a Finite Horizon. Manufacturing & service operations management. [Online] 7 (1), 74–80.

Tomlin, B. (2006) On the Value of Mitigation and Contingency Strategies for Managing Supply Chain Disruption Risks. Management science. [Online] 52 (5), 639–657.

Tseng, S. M. (2014) The impact of knowledge management capabilities and supplier relationship management on corporate performance. International Journal of 18 Production Economics. Elsevier. [Online] 154, 39–47.

Venkatadri, U. et al. (2021) A Model for Demand Planning in Supply Chains with Congestion Effects. Logistics. [Online] 5 (1), 3–.

Voss, C., Tsikriktsis, N. &Frohlich, M. (2002) Case Research in Operations Management, International Journal of Operations and Productions Management. [Online] 22, 2, 195-219.

Xiao, T. et al. (2014) Price and lead time competition, and coordination for make-to-order supply chains. Computers & industrial engineering. [Online] 6823–34.