MASTER

Organizational learning
on the relation between organizational tools and organizational performance

van de Watering, J.L.A.

Award date:
2018

Eindhoven, October 2018

Organizational Learning:

On the Relation between Organizational Tools and Organizational Performance

By J.L.A. (Joey) van de Watering BSc.

Student number: 1027359

In partial fulfillment of the requirements for the degree of

Master of Science

in Innovation Management

Supervisors:

1st Supervisor : Dr. A.A. Alblas, Eindhoven University of Technology, ITEM

2nd Supervisor : Dr. N. Raassens, Eindhoven University of Technology, ITEM

3rd Supervisor : Prof. Dr. F. Langerak, Eindhoven University of Technology, ITEM

ASML Mentor : S. Schepens, ASML Holding N.V.

Eindhoven University of Technology (TU/e)

Faculty of Industrial Engineering & Innovation Sciences

Department of Innovation, Technology, Entrepreneurship & Marketing

Subject headings: learning curve, organizational learning, organizational performance, organizational tools, IT tools, knowledge management, knowledge management processes, knowledge management systems, knowledge repositories.

ACKNOWLEDGEMENTS

This document contains the final result of my master thesis, which I conducted in partial fulfillment of the Master of Science degree in Innovation Management at Eindhoven University of Technology. With finishing this document, an end has not only come to achieving my master degree, but to my life as a student at all. The last six years of being a student have taken me to three different universities, where I have met many people that have contributed to my learning processes in some way, both on a personal and on a professional level. I would like to take a moment to show my gratitude for the ones that have supported me during these years.

First of all, this master thesis could not have been possible without the guidance, criticism and support of the people I worked with in close collaboration. First of all, I would like to thank Dr. A.A. Alblas. Not only has he brought me in touch with the right people to do my graduation project at ASML, but he has also been a great mentor during my time at Eindhoven University of Technology. Through providing support and critical reflections on my decisions to pick certain courses, also during my exchange semester, have I received the right guidance to achieve the most optimal result of my time at this university. Secondly, although all ASML employees have been very willing to help me, I would like to thank my ASML mentor S. Schepens and ASML PhD candidate F. Zijlstra in particular. They have contributed to helping me understand the processes at ASML and making sure I created an understanding of the overwhelming amount of information and data at ASML. Thirdly, fellow ASML graduate intern J. Klaver has been a big source of inspiration for me with techniques for cleaning data, general R commands and models for statistical analysis. Finally, I would like to thank Dr. N. Raassens for providing me thorough feedback on my concept version.

Additionally to the people that have helped me with this project, would I not have been able to achieve the presented result without my friends, family and loved ones, who unconditionally support me and give me the opportunity to distract myself when I need it. As this research is on learning curves; I consider all of you to have been the supporting mechanisms behind my own learning curve and I excitingly look forward to where the future development of my learning curve will bring me to.

MANAGEMENT SUMMARY

This research investigates the effects of an organizational knowledge base, stored in organizational information technology (IT) tools, and its effects on organizational performance. This is done in light of organizational learning theories. Organizational IT tools have previously been identified to function as knowledge repositories, but the effects of a growing knowledge base as a result of organizational learning processes on organizational performance have not been found investigated yet. Additionally, how organizational learning takes places within service industries has received little attention yet. This research tries to fill both research gaps through conducting a longitudinal case-study. It does so by looking at the improving performance of ASML's customer support (CS) department and the most important tools that this department uses to achieve this. The investigated tools are the service diagnostics tool (SDT) and the work instructions (WI), which are used on a global scale to minimize unscheduled downtime, total downtime and costs. Firstly, a theoretical framework is constructed through performing a literature study. Hereafter, this framework is tested through performing a multiple regression analysis. Panel data on five different machine types of the TWINSCAN platform is being used as input for the regression. Formal tests are conducted to decide between pooled OLS model, a fixed effects model and a random effects model. The findings of this research show that the volume of the knowledge base in organizational IT tools significantly contributes to the organizational performance, although this effect is only observed in relation to unscheduled and total downtime and not in relation to costs. More specifically, for every percental volume increase of the SDT knowledge base, the unscheduled downtime decreases with 1.397%. Moreover, for every percental volume increase of the WI knowledge base, the total downtime decreases with 0.436%. Additionally, it is found that the effects of the knowledge base are moderated by the quality of the documents that the knowledge base contains and the validation of these documents. The quality moderation effect shows that the effect of quality diminishes when volume increases, whereas the validation moderation effect shows that the effect of document validation increases when the volume of the knowledge base increases.

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

TABLE 1: OVERVIEW OF THE USED ABBREVIATIONS

Abbreviation	Meaning
CS	Customer support
IT	Information technology
SDT	System diagnostic tool
WI	Work instruction
KMS	Knowledge management system
KRP	Knowledge repository portals
TMS	Transactive memory system
TR	Virtual team rooms
ECOP	Electronic communities of practice
PBTM	Pattern-based task management
ER	Expert rule
PCS	Problem-Cause-Solution
WA	Workaround
SO	Service order
TA	Technical author
FIFO	First-in-first-out
LIFO	Last-in-first-out
WoW	Way-of-Working
CV	Control variable

TABLE OF CONTENTS

1. INTRODUCTION

The handling of information and the management of knowledge has become increasingly important for organizations to achieve a sustainable competitive advantage (Gold, Malthora, & Segars, 2001). This competitive environment in which organizations operate is often called the knowledge economy. In the knowledge economy, organizations wish not to focus on activities that deliver diminishing returns, but wish to focus on activities that deliver increasing returns. The result of focusing on activities that deliver increasing returns is an increasing gap between followers and the leader of the industry, i.e. an increasing competitive advantage (Teece, 1998). To achieve a situation of increasing returns, organizations must develop absorptive capacity; the ability to use prior knowledge to recognize the value of new information, assimilate it and apply it to create new knowledge and capabilities (Gold et al., 2001).

Creating absorptive capacity can be done through the basic elements that form organizations, which are the organizational members, the organization tools and the organizational routines (Argote & Miron-Spektor, 2011). These basic elements are able to form networks through which the different knowledge management processes, knowledge creation, knowledge storage and knowledge transfer take place (Argote & Miron-Spektor, 2011). Analyzing the effects of the knowledge management processes within organizations has only become a mainstream discipline in science since the end of the 1990s (Teece, 1998). Acknowledgement of the importance of developing knowledge assets for organizations has come together in the field of *organizational learning*. This field concerns the science of generating and using organizational knowledge and includes disciplines from sociology, economics, organization sciences, information sciences and engineering sciences (Argote & Miron-Spektor, 2011). Researchers from the field of organizational learning have tried to develop learning curves, which for example show the reduction of production costs over time, the reduction of production failures over time, the reduction of manufacturing time over time or the increasing reliability over time (Argote & Epple, 1990).

Studying the learning curves of different organizations has shown varying results. Variety even exists for organizations operating in the same industry or between geographically dispersed locations of the same organization (Lapré, 2010). Variations in the learning rates can have several reasons, such as organizational forgetting, employee turnover, exogenous factors (e.g. new competitiors) and knowledge transfers (Argote & Epple, 1990). Knowledge transfers take place through the various knowledge management processes of organizations and organizational tools may help to support these processes.

1.1 PROBLEM SETTING

ASML is the market leader for organizations operating in the field of lithography systems for the semiconductor industry. They have an 85% market share (Lex, 2015). Being a market leader brings competitive advantages (Schoeffler, Buzzell, & Heany, 1974). There are several possible reasons for ASML having become the market leader. For example, ASML has been trying to increase the performance of its technology to meet Moore's law. This law states that the density of components per integrated circuit doubles at regular intervals (Schaller, 1997). To achieve this, the organization constantly tries to improve their technology through various learning processes. However, having the right technology is not the only possible explanation for having become the market leader. Vandermerwe & Rada (1988) argued that the organizations of the future with a competitive advantage will be the ones that offer services with their products, as ASML does with their after-sales department 'Customer Support' (CS). Hence, this could be a second reason for their competitive advantage.

Learning processes are expected to apply both to developing technological competencies as they do for service processes. However, the learning rate for services is less well understood within ASML, nor have learning rates related to services been thoroughly investigated. This is addressed by Valtakoski (2017), who tried to fill this literature gap by developing a knowledge-based view framework for service delivery. Understanding the mechanism of learning and after-sales services is interesting, considering the prospected growth of organizations delivering services with their products (Baines & Lightfoot, 2013). For ASML's CS department, it is not known what the driving mechanisms behind the organizational learning processes are and/or to which extent they contribute to an increasing performance.

Clarifying the mechanisms behind these basic elements are important, because resources can be more effectively managed and applied when they are known.

A starting point to search for the answers to these questions may be the organizational information technology (IT) tools of ASML. For large firms that operate in highly competitive environments, such as ASML, these tools will affect organizational knowledge management processes (Alavi & Leidner, 2001). Two reasons for this can be found in the nature of IT tools. Firstly, IT tools can preserve acquired knowledge for an unlimited amount of time. Secondly, IT tools allow the diffusion of the acquired knowledge on geographically dispersed locations through a network connection (Kane & Alavi, 2007; Tippins & Sohi, 2003). As a result of these reasons, organizational IT tools are able to facilitate the fragmented flow of information and knowledge within the boundaries of an organization (Gold et al., 2001). However, while much research has been done on the relation between knowledge management and IT tools, little research has been performed regarding this relationship in light of the organizational learning theories (Ryu, Kim, Chaudhury, & Rao, 2005). Tippins & Sohi (2003) took the concepts of IT competency, organizational learning and organizational performance and suggested that the effects of IT competence on performance were mediated through organizational learning processes. In addition, Ryu et al. (2005) have investigated how experience leads to the development of knowledge stored in tools. However, while these two researches try to fill the gap by looking at the learning processes and IT tools, no research has been found that investigates the relationship between a growing knowledge base in organizational IT tools and the performance of the organization.

This research will try to contribute to both the gap regarding a lack of understanding in learning mechanisms for services and the gap regarding the relationship between knowledge in IT tools and the performance of the organization. It will do so by trying to quantify the contribution of the two mainly used organizational IT tools of ASML that are meant to deliver customer value within the ASML CS department. More specifically, it will investigate the effects of an increasing knowledge base of a tool used for diagnosing machine problems and errors, the *'Service Diagnostics Tool'* (SDT), and a tool meant to support replacing and repairing machine parts, the '*Work Instructions'* (WIs). As will later be discussed more elaborate, the CS department is responsible for minimizing downtime as much as possible. Therefore, the scheduled and unscheduled downtimes will be taken into consideration. Furthermore, the CS department is accountable for the booked costs too, as achieving an optimal performance is preferably done by booking as little costs as possible.

By taking the above into consideration, the main research question that will be answered is: "*how does a growing knowledge base in organizational IT tools contribute to an improved organizational performance?"* Additionally, factors that possibly influence this relationship will be researched through answering: "*what factors influence the relationship between knowledge stored in organizational IT tools and organizational performance?"* To answer these research questions a literature study will be conducted to deepen the understanding on the relation between tools, knowledge and performance. The literature study will be used to construct a theoretical framework. The theoretical framework will be tested through using data of both the SDT and the WIs. The results from this quantitative analysis will be used to draw theoretical and tool specific conclusions that will be used as input for a management advise.

2. LITERATURE REVIEW

Answering the research questions requires an understanding of the processes that lay at the base of generating knowledge, defining the knowledge itself, IT tools and constructs that possibly influence the effects between knowledge stored in IT tools and organizational performance. Creating these insights will be used to develop a theoretical framework.

2.1 ORGANIZATIONAL LEARNING

The general process of organizational learning can be described in different ways. For example, it can be described as the acquisition of information, the dissemination of information, creating a shared interpretation on the information and storing the information in an organizational memory (Tippins & Sohi, 2003). Another way to describe the process is the creation of knowledge, retaining the knowledge and transferring the knowledge (Argote, 2011). Though, what both descriptions have in common is that it is a process of learning that happens over time through the acquisition of experience. In other words, it is the experience that is gathered over time that results in accumulated information and ultimately knowledge (Argote & Miron-Spektor, 2011). Additionally to experience, learning may also take place outside the organizations' boundaries. In that case members, tools or routines can be brought in from outside the boundaries of the organization (Argote & Miron-Spektor, 2011).

TABLE 2: RELEVANT THEORIES IN THE CONTEXT OF ORGANIZATIONAL LEARNING

Sources	*Mechanisms*	*Explanations*
Dorroh et al. (1994)	Learning-by-investment	Learning through a deliberate investment
De Liso et al. (2001), Dorroh et al. (1994) &	Learning-by-doing	Gaining experience by task execution
Ryu et al. (2005)	Learning-from-others	Receive specialized knowledge from other individuals
De Liso et al. (2001)	Learning-by-using	Gaining artifact experience
Alsina et al. (2018)	Machine-learning	Automated data-driven learning

As mentioned earlier, organizational learning does occur through the basic elements organizations are made of, i.e. members, tools and routines, and the networks that can be formed between them (Argote & Miron-Spektor, 2011). Table 2 provides an overview of the theories behind the different ways through which the basic elements of organizations can learn. There is the concept of learning-by-doing, where learning occurs through the accumulation of experience and may result in improved ways of executing a task (De Liso et al., 2001; Dorroh et al., 1994). In addition, there is learning-by-investment, where purposefully resources such as time and effort are used to increase knowledge (Dorroh et al., 1994). Another well-known form of learning is learning-from-others, which enables members of an organization to gather the required knowledge through communicating with other organizational members (Ryu et al., 2005). De Liso, Filatrella & Weaver (2001) mention that next to learning-by-doing, learning-by-using exists. Here, the more an artifact is being used, the more the properties of the artifact are known, which results in a better performance or even improvement of that artifact (De Liso et al., 2001). Lastly, a new form of learning exists, machine-learning. Machine learning is a data-driven form of learning executed completely by software that trains itself through the analysis of that data (Alsina et al., 2018). Because of this feature, one could argue that raw data input is a form of experience. These various learning theories are the starting point of organizational learning and resultingly organizational knowledge and are described to understand the origins of the organizational knowledge. In defining the theoretical framework will the learning processes be left out of scope, as decided upon in the previous section.

2.2 ORGANIZATIONAL KNOWLEDGE

As said, the previously discussed types of learning are expected to generate knowledge. However, knowledge itself is not an unambiguous term and various definitions of it exist. One commonly held believe is the distinction between tacit knowledge and explicit knowledge. Most often explicit knowledge is seen as the knowledge that has been articulated and codified to make it possible to be transferred (Nonaka & von Krogh, 2009). On the contrary, tacit knowledge is seen as unarticulated and uncodified knowledge, which is tied to the senses, entails movement skills, and may consist of physical experiences (Nonaka & von Krogh, 2009).

The nature of these characteristics make tacit knowledge to be believed difficult to transfer (Argote, Mcevily, & Reagans, 2003). However, a second stream of researchers view tacit and explicit knowledge as two types of knowledge that depend on each other. They believe that both types of knowledge exist along a continuum and that they cannot exist without the other (Nonaka & von Krogh, 2009; Tsoukas, 1996). This continuum allows for the conversion from explicit knowledge to tacit knowledge and vice versa (Nonaka & von Krogh, 2009). As this research will focus on the knowledge stored in tools and tools contain documents, inherently explicit knowledge is treated. However, it should not be forgotten that the possible interpretation of these codified documents may dependent upon the tacit component in the mind of the knowledge recipient.

TABLE 3: PRACTICAL KNOWLEDGE CLASSIFICATION THEORIES

Authors	*Classification*	*Definitions*
Brown & Duguid (2001)	Sticky knowledge	Knowledge that is difficult to transfer
	Leaky knowledge	Knowledge that transfers (too) easily
Ryu et al. (2005)	Depth of knowledge	Task specific knowledge
	Breadth of knowledge	The diversity of knowledge
Alavi & Leidner (2001)	Know-what (Declarative knowledge)	Knowledge about something
	Know-how (Procedural knowledge)	Knowledge how something works
	Know-why (Causal knowledge)	Knowledge why something happens
	Know-when (Conditional knowledge)	Knowledge when something happens
	Know-with (Relational knowledge)	Knowledge of the interactions

In addition to the tacit and explicit knowledge classification, more practical definitions and classifications of knowledge do exist, of which an overview is presented in Table 3. One of them is the division of knowledge between sticky and leaky knowledge (Brown & Duguid, 2001). Sticky knowledge is considered to be knowledge that is difficult to transfer, even within the boundaries of the organization. It is to some extent similar to the aforementioned tacit knowledge. Leaky knowledge, on the other hand, is knowledge that is (too) easily transferred within and between organizations. The distinction between sticky and leaky knowledge is one that is often associated with know-how knowledge for sticky knowledge, and know-what knowledge for leaky knowledge (Brown & Duguid, 2001). Know-how knowledge can be associated with knowledge-depth, as a deeper understanding of tasks increases the execution speed of tasks. Know-what knowledge can be associated with the breadth of knowledge, as a wider understanding gives a better understanding of what is happening in a process (Ryu et al., 2005). Alavi & Leidner (2001) name know-how as procedural knowledge and name know-what as declarative knowledge. Additionally, they have added three more categories. Firstly, know-when, which is conditional knowledge, secondly know-why, which is causal knowledge, and thirdly, know-with, which is relational knowledge. This five-category knowledge classification system is practical and allows further categorization of knowledge into documents that are useful for the organization and knowledge documents that are not, which is the pragmatic view on knowledge. This pragmatic approach tells that any information stored by an organization is useful to it (Alavi & Leidner, 2001). Therefore, it can be argued that any type of information stored in organizational IT tools is of use to the organization. To better understand this, in the following section different types of IT tools and the information that they contain will be discussed.

2.3 ORGANIZATIONAL IT TOOLS

As mentioned earlier, IT tools in large organizations are expected to affect organizational knowledge management processes (Alavi & Leidner, 2001). This is because research suggests that IT tools have the ability to support the underlying processes of knowledge sharing and the creation of an organizational memory (Kane & Alavi, 2007). Other proof of the importance of organizational IT tools for organizational learning was found by Kane & Alavi (2007), who found that organizations may positively be affected by the introduction of IT tools to support the organizational learning mechanisms, if they are introduced under the right circumstances.

TABLE 4: OVERVIEW OF THE IT TOOLS FOUND IN THE LITERATURE

Authors	*IT Tools*	*Explanations*	*Document production*	*Knowledge Types*
Kane & Alavi (2007)	Knowledge repository portals	Documentation storage environment	Coded-best practices	Know-about, Know-how & know-with
			Corportate directories	All
			Knowledge networks / Transactive memory system	Know-about

	Team Rooms	Virtual meeting environment	Meeting history	Know-about, Know-with
	Electronic communities of practice	Virtual communication environment	(e.g.) E-mail & chat logfiles	Know-about, Know-with
Kimmerle et al. (2010)	Social-tagging systems	To tag organizational resources	Tags	Know-about
	Pattern-based task management	The documentation of processes	Workflows & improvement suggesstions	Know-what, Know-how & know-with
	Wikis	To document and share facts	Wiki articles	Know-about

Various types of IT tools exist that positively may contribute to knowledge management processes and an overview of them is shown in Table 4. In general, there are organizational IT tools that support organizational learning through connecting individuals and social structures (Kimmerle et al., 2010) and through acting as a knowledge management system (KMS) (Alavi & Leidner, 2001). A type of KMS are knowledge repository portals (KRP), which are meant to store and retain knowledge within the boundaries of the organization (Kane & Alavi, 2007). KRP may contain various types of documents such as coded best practices and corporate directories. Corporate directories map the internal expertise and experience to help find the right expertise at a later point in time (Alavi & Leidner, 2001). Additionally, KRP may contain knowledge networks that should help to find the right relevant knowledge within organizations (Alavi & Leidner, 2001). These knowledge networks have also been called transactive memory systems (TMS) (Wang, Huang, Davison, & Yang, 2018).

As knowledge building takes place in socio-cultural environments, IT tools that enhance social interactions are expected to be beneficial to the knowledge management processes. Kane & Alavi (2007) discuss two types of IT tools. Firstly, virtual team rooms (TR) allow teams to discuss team specific issues without the need to travel. Secondly, electronic communities of practice (ECOP) can be used for communication matters. ECOP consists of various communication technologies, such as e-mail and instant messaging and as a result the documentation types may vary (Kane & Alavi, 2007). Additionally, Kimmerle et al. (2010) discuss three more types of IT tools that support organizational learning and knowledge creation through social interactions. These are (1) social-tagging systems, (2) pattern-based task-management (PBTM) systems and (3) wikis. The social tagging system allows members of an organization to tag various resources throughout the organization, which results in metadata of that resource. The information in the metadata may lead to new concepts and modifications of an individual's cognitive structure. Naturally, the more users use the tagging system, the more valuable the metadata becomes (Kimmerle et al., 2010). PBTM systems are a type of workflow management systems. These systems support members of an organization in carrying out routine tasks, much like the earlier mentioned coded best practices. Workflows are usually made in top-down structure, but the more complex the task, the more flexibility is required to do accordingly, as more personal expertise is required (Kimmerle et al., 2010). PBTM allows members to create workflows bottom-up and to store them in a shared repository, so that different patterns of task execution are available to the entire organization. Users can modify existing patterns to optimize the process in several iterations (Kimmerle et al., 2010). PBTM mainly concerns procedural knowledge (Kimmerle et al., 2010). Lastly, wikis are webpages that contain information on various topics. These webpages can be adjusted by different members of the organization and a log of the changes is kept. By doing so, wikis represent the knowledge of the collective that has access to it. Wikis primarily concern declarative knowledge (Kimmerle et al., 2010). The last column of Table 4 shows the types of knowledge that the documents may consist of, based on own insight and information in the discussed papers. Using this table should help to identify and understand what types of tools the SDT and WIs are and what types of knowledge they contain.

2.4 ORGANIZATIONAL PERFORMANCE

The different types of knowledge stored in the different types of documents in different types of organizational IT tools are expected to contribute to organizational performance. Researchers that investigate learning curves often define organizational performance from a behavioral perspective. This means that as a result of learning and an increasing amount of knowledge, changes are expected in either behavior or characteristics of performance, such as speed or accurateness (Argote, 2013). Contrarily, it may be possible for organizations to gather knowledge and not change behavior or characteristics of performance (Argote, 2013). As the pragmatic view on organizational knowledge allows to declare all documents that an organization possesses are of value to the organization, all documents stored in organizational IT tools are valuable. This is in line with the proposition of Haas & Hansen (2005), who argue that the development of the number of documents stored in a knowledge management tool is a proper knowledge asset measure, with the more documents, the more knowledge. Hence, it is possible to measure the knowledge level of an IT tool over time by the number of documents stored in that IT tool. The number of documents stored at any time can thus be described as the size of the knowledge base. By taking a behavioral perspective approach, a change in behavior or characteristics of the performance of the organization should be noticable as a result of a growing knowledge base. As a result, the following hypothesis regarding the relationship between knowledge and performance can be derived:

Hypothesis 1: The bigger the knowledge base of an organizational IT tool in a given timeframe, the better the organizational performance will be in the following timeframe.

2.5 INFLUENCING CONSTRUCTS

TABLE 5: OVERVIEW OF LITERATURE ON DOCUMENT QUALITY

Author	*Construct*	*Dimension*
Nelson et al. (2005)	Quality	Accuracy
		Completeness
		Currency
		Format
Fadel et al. (2009)	Validation	Content ratings
		Review analysis
Devaraj & Kohli (2003)	Utilization	Performance metric dependent

Having organizational IT tools in place that contain documents is not the only pre-requisite for successful knowledge management processes within organizations and organizational performance as a result. An overview of constructs that possibly influence this relation are shown in Table 5. Firstly, the quality of the documents that an organizational IT tool contains is a pre-requisite for a successful organizational IT system (Nelson et al., 2005). Nonetheless, when the documents contain (e.g.) incorrect or limited information, would it be logically to expect that the performance of this tool is lower than when it would contain purely correct and complete information. The quality characteristic of documents can be divided into four dimensions, i.e. accuracy, completeness, currency, and format (Nelson et al., 2005). In corresponding order, they can be described more elaborate as the correctness of information, the degree to which all possible relevant information is stored, the level of information being up-to-date, and the degree to which information is presented in an understandable and interpretable manner to the recipient (Nelson et al., 2005). Therefore, the second hypothesis is:

Hypothesis 2: The higher the quality level of the knowledge base of an organizational IT tool in a given timeframe at any volume, the better the organizational performance will be in the following timeframe.

Fadel et al. (2009) propose that next to quality, the validation of information is an important determinant for the performance of an organizational IT tool. According to their research, content validation is key to avoid the buildup of low quality knowledge, i.e. incomplete, incorrect and/or obsolete information (Fadel et al., 2009). Knowledge validation mechanisms like content ratings by users themselves (Poston & Speier, 2005) or review analyses by committees have been proposed as a solution (Marwick, 2001). In the content rating validation mechanism, documents are given a score for the quality that it contains, which can be done by users or experts. This rating is inherently subjective and may lead to a mismatch between true and perceived quality (Poston & Speier, 2005). The review analysis committee are quality judgements by experts and the quality judgement may even function as a gateway for distribution, such as with scientific journals (Marwick, 2001). As validation of knowledge has indeed shown to have a positive effect on the perception of knowledge usefulness (Fadel et al., 2009), the third hypothesis is the following:

Hypothesis 3: The higher the validation level of the knowledge base of an organizational IT tool in a given timeframe at any volume, the better the organizational performance will be in the following timeframe.

Thirdly, the actual usage of an organizational IT tool is expected to be important. Nonetheless, an organization can have as many organizational IT tools in place, but if they are not used they are not contributing to the organizational performance. Hence, users should actually use the technology. However, often in IT literature the actual usage of IT systems is overlooked (Devaraj & Kohli, 2003). To measure the impact of the use of IT systems, it is important that the metrics of the actual usage of IT systems are tied to the metrics of the organizational performance (Devaraj & Kohli, 2003). As actual usage was found in a different longitudinal case-study to be significantly and positively correlated with performance (Devaraj & Kohli, 2003), does this research elaborate on that by hypothesizing that:

Hypothesis 4: The higher the usage rate of the knowledge base of an organizational IT tool in a given timeframe at any volume, the better the organizational performance will be in the following timeframe.

2.6 PROPOSED THEORETICAL FRAMEWORK

Based on the hypotheses described above, a generic framework for knowledge buildup in tools and the expected effect of it on organizational performance can be drawn. The result of this is shown in Figure 1. This proposed framework will be tested in the following sections through using data of both the aforementioned organizational IT tools at ASML, the SDT and WIs. First these tools will be investigated more elaborately, after which the variable operationalizations are discussed and used as input to create empirical models. These will be tested to find support for this model.

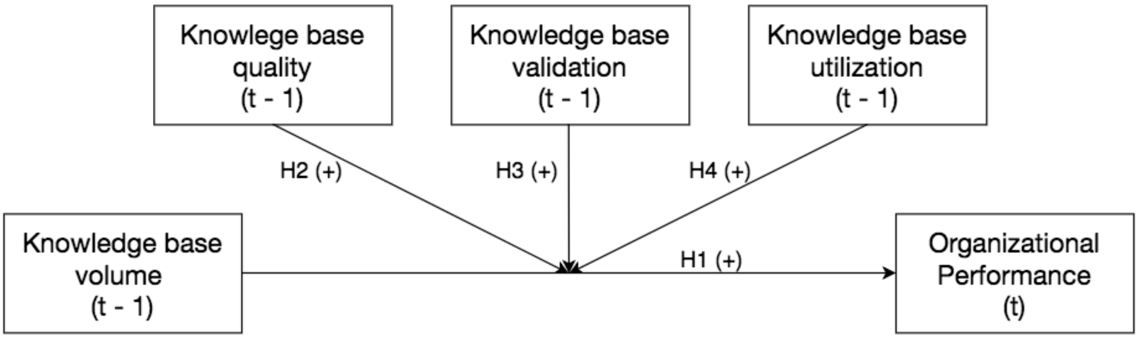

FIGURE 1: PROPOSED THEORETICAL FRAMEWORK

The research questions will be answered in a quantitative matter by performing multiple regression analyses to test the hypotheses of the proposed theoretical framework. A multiple regression analysis is a widely used tool to predict a dependent variable by using one or more independent variables. The proposed theoretical framework will only be tested on one machine platform, the TWINSCAN platform, due to resource limitations. This platform contains five different machine types that were introduced sequentially; the NXT:1950i, the NXT:1960Bi, The NXT:1965Ci, The NXT:1970Ci and the NXT:1980Di. The sequential introduction and installed base of these machine types is shown in Figure 2.

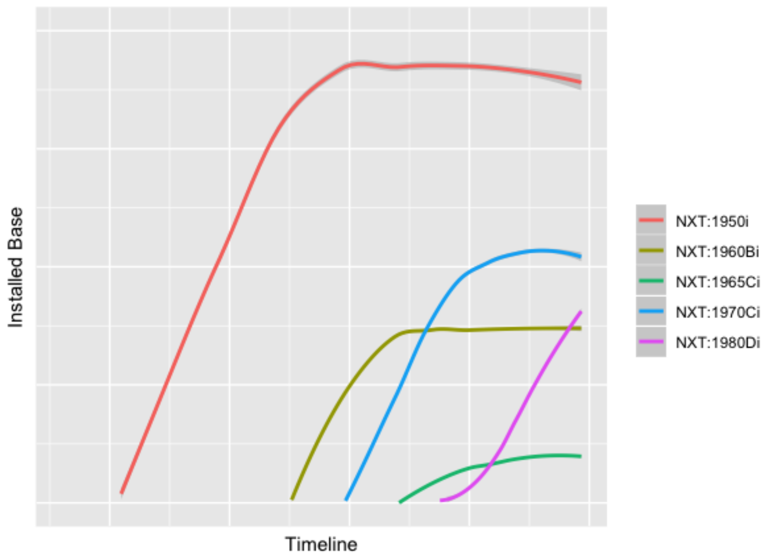

FIGURE 2: THE INSTALLED BASE OF THE NXT MACHINES OVER TIME

Learning curves most often depict a type of performance on the y-axis and a type of experience on the *x*-axis to show the performance development over time (Argote & Epple, 1990). Classically, the learning curve can econometrically be described as in Equation 1 (Argote, 2013):

$$y_i = ax_i^b$$

EQUATION 1: THE CLASSICIAL LEARNING CURVE FORMULA

- y = the number of labor hours per unit
- a = the number of labor hours to produce the first unit
- x = the cumulative units produced in a timeframe i
- ^{b} = the learning rate
- $_i$ = the timeframe

An adjusted version of this formula can be used for developing an understanding of the reasons behind the existence of a learning curve at the ASML CS department and the effects of it on organizational performance. In the following section will first the empirical setting be described, after which the tool selection is discussed together with the variable operationalizations.

3.1 EMPIRICAL SETTING

Firstly, the empirical setting of this research must be explained to make Equation 1 applicable to this research. This is done in the following section where first the performance metrics are defined. Thereafter, the tools, their purpose and the tool metrics will be discussed.

3.1.1 THE SERVICE PERFORMANCE

ASML provides the opportunity to close a service contract when selling an NXT machine to the customer. When a customer decides to close a service contract, ASML becomes responsible for the availability of the machines at the location of the customer. ASML uses the machine state definitions from the global industry association, the Semiconductor Equipment and Materials International (SEMI) to do so. These standards have been approved by the Global Metrics Committee (SEMI, 2004). The various states that a machine can be in are shown in Figure 3.

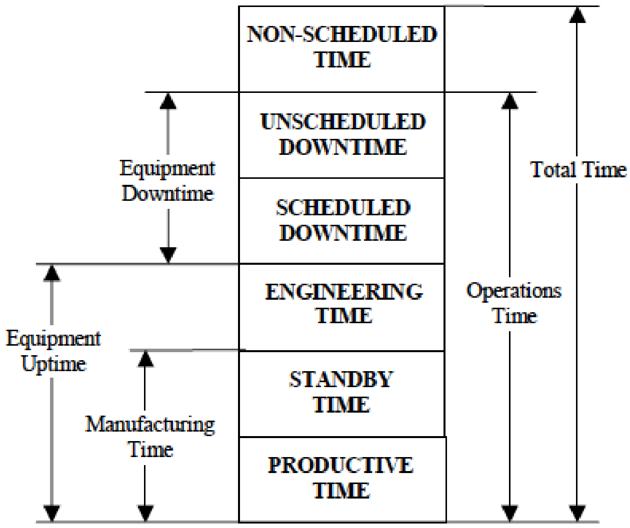

FIGURE 3: SEMI MACHINE STATES

Not every state is considered in calculating the machine availability. To calculate machine availability, ASML uses the formula as presented in Equation 2 (SEMI, 2004).

$$Availability = \frac{Equipment\ uptime\ x\ 100}{(Operations\ Time - (Non\ ASML\ Accountable\ Downtime))}$$

EQUATION 2: FORMULA TO CALCULATE MACHINE AVAILABILITY

To achieve an optimal availability percentage, the CS department ensures that when a machine is down it is restored to its initial state as soon as possible. Hence, the CS department tries to minimize downtimes as much as possible. In case a machine experiences a down that was caused by the customer, then that down is not considered for calculating the availability percentage. This is because ASML is not accountable for those times in its service contracts. However, customers often request the help from ASML for those cases that the customer is liable for the down, because ASML possesses the knowledge to restore the machine as quickly as possible. ASML then helps the customer by using their own knowledge and tools. The parts and labor needed to restore a machine to its initial state during a customer liable down are billed to the customer.

Both ASML accountable and non-accountable downs are either scheduled or unscheduled. SEMI (2004) defines scheduled down as the time when the equipment is not available to perform its intended function due to planned downtime events, which include maintenance delays, production tests, preventive maintenance, change of consumables or chemicals, setup and facilities related downs (SEMI, 2004). Unscheduled downtime is defined as the time when the equipment is not in a condition to perform its intended function due to unplanned down events, which include maintenance delays, repairs, changes of consumables or chemicals, out-of-spec-input and facilities related downs (SEMI, 2004).

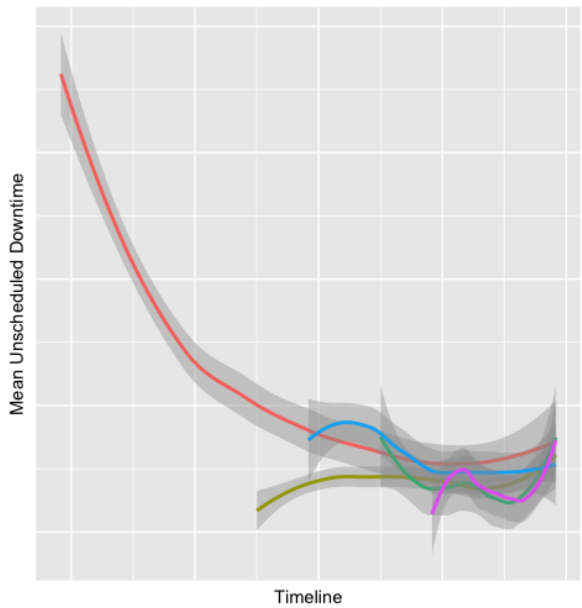

FIGURE 4: DEVELOPMENT OF THE UNSCHEDULED DOWNTIME CURVE

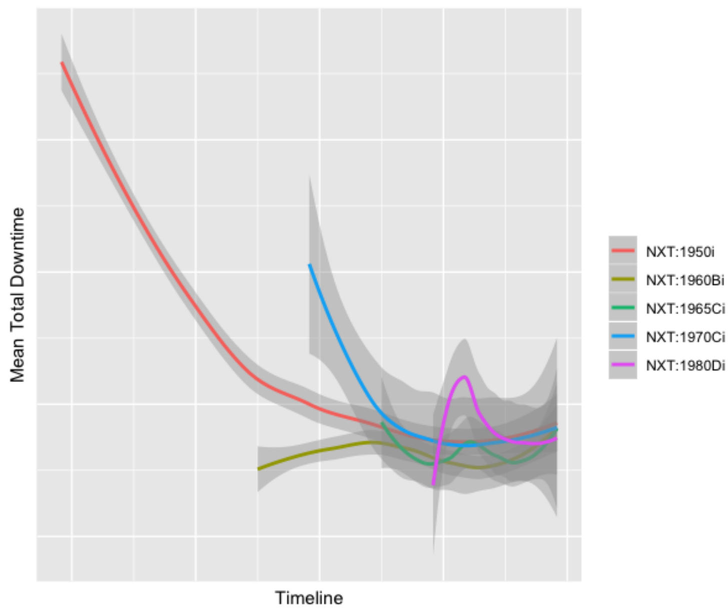

FIGURE 5: DEVELOPMENT OF THE TOTAL DOWNTIME CURVE

ASML has various tools to support the CS engineers in bringing a machine back to work. The tools that are most frequently used by CS to achieve this are the *Service Diagnostics Tool* (SDT) and the *Work Instructions* (WIs). The SDT is mainly used during the unscheduled downtime, whereas the WIs are used during both scheduled and unscheduled downtime states. When a machine is brought back to its initial state, several start-up tests and calibrations must be performed, which is known as the recovery state. It is expected that the knowledge bases of the SDT and WIs contribute to lower downtimes. Hence, since ASML engineers and the supportive tools are used during both ASML accountable and non-accountable downs, and the primary concern of the engineers is to bring the machine back to its initial state as quickly as possible, the unscheduled downtime for the SDT and the scheduled downtime for the WIs will be used as performance measures for this research. These developments of the two performance variables are depicted in Figure 4 and Figure 5.

Additionally, the costs to achieve minimal downtimes will be taken into consideration too. Costs can be expressed in both labor and parts. There are various reasons to believe that knowledge in tools lead to lower costs. First of all, having knowledge lowers the occurences for which engineers perform the wrong actions or order the wrong parts. This in turn is expected to lower the quantity of labor hours and decrease costs when ordering parts. For this research, the effects of a growing knowledge base on the mean booked costs of a certain timeframe will be analyzed. The development of the mean booked costs over time is shown in Figure 6.

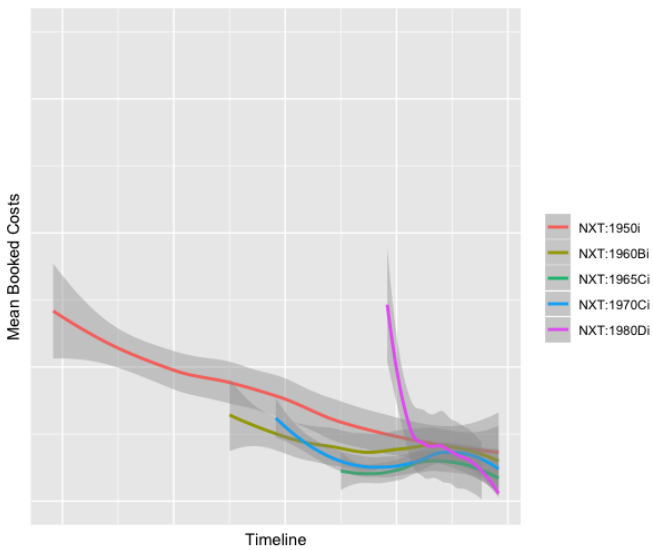

FIGURE 6: MEAN BOOKED COSTS DEVELOPMENT OVER TIME

Information and data on the downtimes and costs are retrieved from ASML's SAP databases. Retrieving the information from SAP has resulted in two separated databases:

1. PMA: a database containing various machine performance measures per machine
2. CSCA: a database containing information on the booked costs per machine

3.1.2 THE SERVICE DIAGNOSTICS TOOL

The first tool that is expected to contribute to this increased performance is the SDT. The SDT is used by engineers to analyze the problem of an unscheduled down from the moment that the down occurs. The SDT is a KRP in which all information stored can be accessed through a stable and secured internet connection. The information in the SDT is a historical list of all problems with their causes and their possible solutions. The learning and experience processes that fill the knowledge in this tool are a collaboration between learning-by-others and machine learning.

In case a machine encounters a problem, a list of unique error codes is send out to the SDT servers. This unique list of error codes is often called a *'fingerprint'.* All problems have a fingerprint that is stored along with them within the SDT. Whenever a machine somewhere in the world encounters a problem, that fingerprint can be matched to all the fingerprints that are stored within the system by an algorithm that runs on the background. The algorithm presents an estimation of the likelihood that fingerprint x is similar to fingerprint y. Hence, the performance of the SDT is dependent upon the amount of information it contains, with the more problems and fingerprints, the higher the likelihood of a possible match and accurateness of the presented match. Nonetheless, when it contains no information, no estimation can be given. However, the same problem can have multiple lengthy fingerprints that only slightly differ. In addition to the fingerprints that are put in the SDT through machine failures, there are fingerprints that are put in the SDT through personal expertise of an engineer. These fingerprints are called '*expert rules'* (ERs). ERs are problems with a small list of error codes that are always contained in fingerprints for those problems. Essentially, the ERs list the key error codes for a problem. Because of this, the algorithm values ERs higher in the problem matching process than lengthier fingerprints. A problem with the ERs arises when the error codes are too generic. In that case the ER results in too many matches.

More fingerprints should contribute to more accurate matches to find the right solution to the encountered problem, if that solution is known. Solutions are written and added to the SDT by the engineers that solve these problems. As not necessarily all problems with a fingerprint are serious problems, not all do and will have a solution stored with them. Some fingerprints may be the result of a hiccup in the system, which do not have to be resolved. All fingerprints that do not receive a corresponding solution are removed from the SDT three months after the occurrence of that problem. However, when a fingerprint is the result of a more serious problem, a solution for that fingerprint can or must be stored in the SDT. If a problem results in an unscheduled down of > 3 hours, a solution must be added. In all other cases a solution can be added. Adding a solution to a problem and cause (the fingerprint) makes it a Problem-Cause-Solution (PCS). To create a PCS, all aspects of the problem, including the causes and the solution to resolve the issue, must be well known. For cases where these aspects are less well known can a second type of solution, a workaround (WA), be written. In a WA the reasoning behind the solution is less well known but a method to resolving the issue is known through executing certain steps. Sometimes these steps are as simply as restarting the drivers. Generally, PCSs are considered to be of a higher quality than WAs. Sometimes a problem and cause can have both a PCS and a WA. In such cases, the solution may be to do the WA first and then do the PCS.

While the SDT system is there to support the engineers in analyzing the problem to resolve the unscheduled down as swiftly as possible, the usage is not always mandatory. Only for the unscheduled downs that take > 2 hours to resolve, the use of it is mandatory. However, all use of the SDT is being traced. When a user of the system clicks on the first given solution in the SDT, a timestamp is send out to a database. It is assumable that resolving the problem that caused the unscheduled down without the SDT takes more time than when using the SDT. In addition, when the SDT is used as soon as possible after the occurrence of the problem, the possible solution is presented as soon as possible. Hence, the machine will be up as soon as possible.

Based on the above information, the databases that are available for analysis and having multiple in-depth interviews with engineers of the SDT, result in using the following data and information:

1. SDT: a database that contains all problems, fingerprints and links to solutions
2. ER: a database that contains all expert rules
3. UGE: a database that contains information on actual SDT usage in relation to an unscheduled down.

Furthermore, a small database exists on the percentage of problems that the algorithm could link to a stored fingerprint and the percentage of occurred fingerprints for which a solution was added to the SDT. This database only shows the average match rate among all TWINSCAN machines as a percentage per month. Furthermore, the measuring methodology has changed halfway the database. Given this, it was decided not to use this database for further analyses as the reliability of the information and completeness of this database is low.

3.1.3 THE SDT EMPIRICAL FRAMEWORK

To test the proposed theoretical framework as presented in Figure 1 must the elements of this figure be filled to make it applicable to this tool specifically, as is done in Figure 7. To test Hypothesis 1, the decision is made to cumulatively count the number of unique problem-fingerprint combinations. This is a way of counting the cumulative number of documents as is proposed by Haas & Hansen (2005) as a knowledge measure. This metric for counting documents is chosen because it can be assumed that the more problem-fingerprint combinations with a solution are stored in the SDT, the higher the likelihood will be that a given problem can be matched to an earlier occurred problem for which a solution was found at some point in time. Hence, the SDT has more knowledge. Additionally, since PCSs are considered to be of a higher quality than WAs, this information is being used for determining the quality of the SDT knowledge base. As said earlier, this is because PCSs are better understood than WAs. This by itself can be traced back to the accuracy and completeness dimensions of the theory on document quality by Nelson et al. (2005). Finally, utilization data is available for the SDT in terms of time to use the SDT. It is decided to use this metric because when lowering unscheduled downtimes, time is crucial and measuring actual usage is organizational performance metric dependent, as was argued by Devaraj & Kohli (2003). Data concerning validation of proposed solutions is not taken into consideration, because the quality of this is low, as mentioned above. Therefore, Hypothesis 3 will not be tested in light of the SDT.

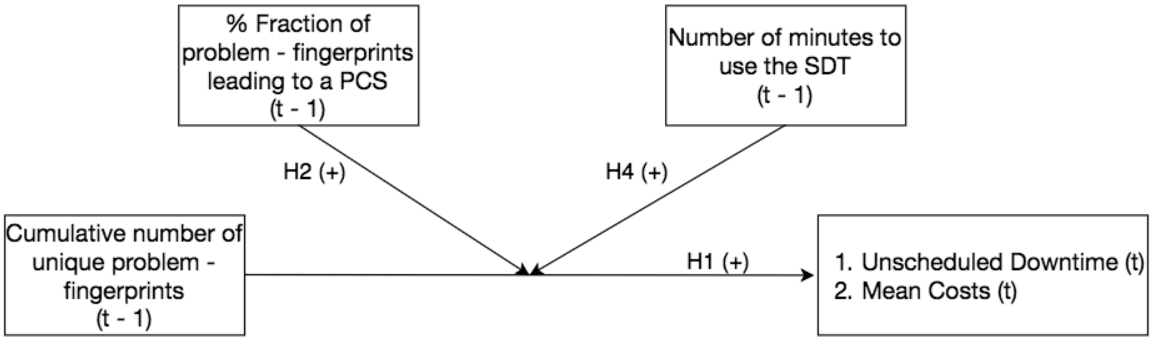

FIGURE 7: EMPIRICAL MODEL FOR THE SERVICE DIAGNOSTICS TOOL

3.1.4 THE WORK INSTRUCTIONS

ASML's WIs contain the procedures to replace and repair parts of a machine. This can happen after the diagnosis of the problem has taken place and a part has to be replaced as a result of it. Alternatively, it can be during a planned event. WIs are similar to what Alavi & Leidner (2001) name the coding of best practices, or what Kimmerle et al. (2010) call PBTM. Therefore, the WIs are identified to contain knowledge on know-what, know-how and know-with. The main learning mechanisms for creating the WIs is learning-by-others and learning-by-investment.

If a part of a machine must be replaced, an engineer creates a service order (SO). This SO tells which actions must be performed and which parts should be used. Each part has a unique 12-digit number, called the 12NC number. The 12NC number links each part to a specific WI that can be used to install that part. The last digit of the 12NC number is updated when a part receives an update. This number is then also updated in the WIs. With each SO and part order there are corresponding costs. Sometimes, when a new SO is created, the engineer does not know yet which parts are needed. Therefore, the engineer orders all possible parts, because delivery times of parts may be long and time is crucial. In case a part is not used, that part is send back and can be found as a negative cost in the cost database CSCA. Other possibilities for SOs with negative costs may be when a part needs replacement but remains high in value. In such cases the part is send back to ASML's supplier so that it can be restored or repaired. A single WI can be used both to take a part out of a machine and to install a part in a machine.

WIs can be modified over time, just like the modifications of the PBTMs as described by Kimmerle et al. (2010). These modifications are often the result of comments placed on that specific WI by the engineers who use the WIs. Two assumptions can be made regarding the comments placed on a WI. Firstly, the specific WI is being used. Secondly, the quality of the WI is still sub-optimal if not all comments have been solved. Comments can be rejected, if they are considered false. In case they are considered to be true they are tried to be solved as soon as possible. However, due to a backlog on the processing of the comments, ASML has switched from a first-in-first-out (FIFO) approach to a last-in-first-out (LIFO) approach. Additionally, a comment can have three different natures. They can be typographical, configurational or technical. Typographical comments concern the language of the WIs, configurational comments concern the structure of the WI and technical comments concern the content information. After processing a comment, a new version of a WI may be released in a global KRP, where engineers can find all of the newest WIs. New versions can either have had a minor or a major update. Although all current versions of a WI are easily accessible in the KRP that engineers use to extract them from, the historical versions of a WI are much more difficult to obtain. The historical versions are (partly) stored in a different KRP. ASML does not know how to extract the content of these historical versions in batches.

Furthermore, WIs can be validated by a technical author (TA). A TA ensures that the information in a WI is in the right format, the correct language is being used and corrects mistakes in the content. In this sense, a select group of TAs operate much like the proposed review analysis committee of Marwick (2001). However, the validation of TAs does not function as a gateway for distribution at ASML. After an update of a WI, the type of author of a WI can have changed from an engineer to a TA or vice versa, even for minor updates. However, it can be assumed that after a minor update the WI has not changed radically and therefore the WI remains validated. This is true until a major update of a WI comes out. A major update is able to completely renew a WI.

It may be possible that production sites use old versions of a WI. This is because the clean-rooms, where work on these machines is performed, make the machines vulnerable to loss of intellectual property. For this reason, engineers are not allowed to bring in internet connected devices. The only device that can be brought in is a '*Fabtop*', which is a stripped laptop that can only create a connection to download the newest version of a WI. These computers are only able to download the newest WIs every 24 to 48 hours. Some production locations have chosen to work with WIs on offline laptops. Offline laptops are normally updated every one to three weeks. For this reason, actual data on the usage of WIs is not being traced. In addition, ASML customers may switch from one way-of-working (WoW) to another WoW with the WIs. Although there is a database that contains the current WoW of all customers, no historical database exists that shows the changes in WoW of customers over time.

By taking the above information into consideration and having in-depth interviews with employees working with the WIs, the following datasets have been identified to be of use to execute this research:

1. CSCA: a dataset containing all booked costs with SOs and used materials
2. NC: a dataset containing all 12NC numbers and their corresponding WIs
3. PCD: a dataset containing the release dates of all historical versions of the WIs
4. CMDS: a dataset containing all processed comments
5. CMDS.un: a dataset containing all comments that have not been processed yet
6. TA: a database that contains the names of all TAs that validate the WIs[1]

3.1.5 THE WI EMPIRICAL FRAMEWORK

Similarly to the SDT, the possible obtainable information from the datasets is combined with the theoretical framework of Figure 1 to create an empirical framework specifically for the WIs, as is shown in Figure 8. Furthermore, like the SDT, the main variable of interest is the growing knowledge base over time and its effects on the organizational performance. To measure this, the number of unique available WIs on a machine type will be taken as the knowledge base variable. This is done as it is possible to assume that more available WIs on a machine type will result in a better performance of these machines through decreasing downtimes, since more knowledge is available. In a similar way, Haas & Hansen (2005) argue that measuring

[1] This database was created specifically for this research as it did not exist before.

documents is an adequate method to measure knowledge in an organizational IT tool. Additionally, the other databases allow for the derivation of all proposed moderating variables. Firstly, the proposed quality moderation effect can be derived from the comments. The nature of the comments, i.e. whether they are typographical, configurational or technical, can be related to the quality dimensions as described by Nelson et al. (2005). More specifically, typographical and configurational comments can be related to increasing the format of the WIs, as they are made to try and improve the language, the order of steps and the number of steps that are needed to correctly handle a part. Additionally, technical comments can be related to increasing the accuracy, currency and completeness of the WIs. Furthermore, it is possible to link the TA validation process to the review analyses, as described by Marwick (2001). This allows to use this information to determine whether a document was validated or not. Finally, the actual usage of it can be determined through assuming that a WI is always used when a version of it is available at the time of performing an action, as is expected through ASML policy. By doing so, this element of the theoretical framework becomes context specific, as is proposed by Devaraj & Kohli (2003).

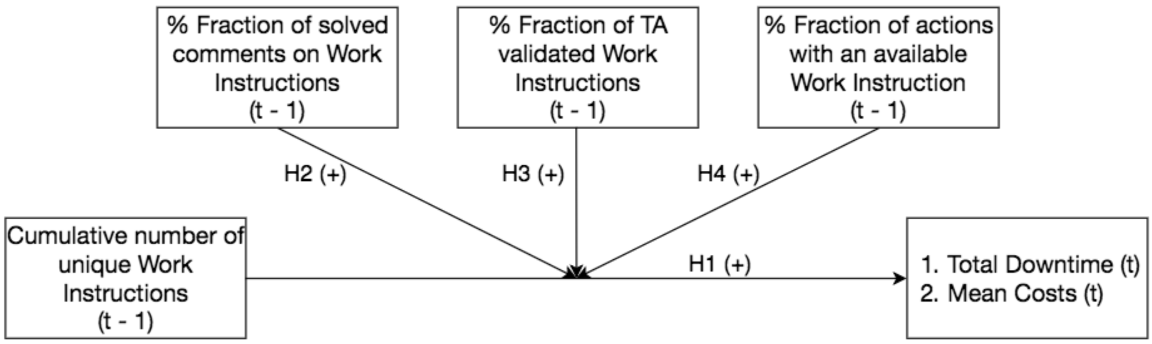

FIGURE 8: EMPIRICAL MODEL FOR THE WORK INSTRUCTIONS

3.2 THE STATISTICAL ANALYSES

To find support for the proposed theoretical framework as presented in Figure 1, statistical analyses of the empirical models, as shown in Figure 7 and Figure 8, will be performed. To do accordingly, multiple regression analyses will be performed using the open-source data analysis software R Studio. A multiple regression analysis is a widely used tool to predict a dependent variable through using one or more independent variables. Panel data will be used to test the hypotheses and answer the research questions. Panel data is a cross-sectioned database with a time component (Wooldridge, 2015). For the database used in this research, the cross-section is made on the different machine types of the TWINSCAN platform, which are the aforementioned NXT:1950i, NXT:1960Bi, NXT:1965Ci, NXT:1970Ci and NXT:1980Di machines. A machine level cross-section is made, because the SDT and WIs can be accessed and used worldwide through an internet connection. Hence, all knowledge is available to every machine of the same type. Panel data also has a time series component. The interval of the time series points is set on a monthly scale. A monthly scale was chosen to filter out the effects between the possible usage of the newest versions of WIs and the actual used versions of the different WIs. By taking a month level, it is assumed that the error within all months is equal. Because of a loss of available data points for analysis, taking a higher scale level, such as years or quarters, is undesirable. To convert to monthly data, the initial data must be converted from a day or week scale. The general multivariate linear regression model for cross-sectional data in a time series with a moderating effect is given in Equation 3, which is adjusted from Greene (2003):

$$y_{it} = \alpha + \beta_1 x_{it} + \beta_2 x_{it} + \beta_3 x_{it} z_{it} + \ldots + \beta_k x_{it} + v_{it}$$

EQUATION 3: THE GENERAL REGRESSION MODEL FOR PANEL-DATA

- Y_{it} = the dependent variable
- X_{it} = the independent variable
- z_{it} = the moderating variable
- $_i$ = the cross section
- $_t$ = the time range of the data
- v_{it} = the composite error

Furthermore, Equation 4 shows the two different components of the composite error:

$$v_{it} = a_i + u_{it}$$

EQUATION 4: BREAKDOWN OF THE COMPOSITE ERROR

- a_i = the unobserved heterogeneity
- u_{it} = the idiosyncratic error term

3.2.1 VARIABLE OPERATIONALIZATIONS

The empirical models of Figure 7 and Figure 8 can be written down in econometric models. To do accordingly, the variables to put in the econometric model must be operationalized. The operationalization of the dependent variables can be obtained from Table 6. According to Argote (2013), control variables (CVs) should be included when assessing knowledge by measuring changes in practices or performance. This is done to account for alternative factors that are not a result of organizational learning, such as material improvements. The CVs are added based on insights from ASML experts and own insights. A variable time is added to capture the effects of an increased performance as a result of simply time. The number of unique customers is added, because each new customer uses a different WoW to deal with downtimes. Lastly, the installed base of all machine types is added, as it is expected that having installed more machines results in more experience and knowledge, and having a better performance as a result. These CVs and their operationalization are also shown in Table 6.

TABLE 6: OVERVIEW OF THE OPERATIONALIZATION OF THE DEPENDENT VARIABLES

Variable Description	*Sort of Variable*	*Variable Symbol*	*Measured as*	*Variable Operationalization*	*Used Database*
Time	Control	TD_{it}	The cumulative number of days since the first data point of machine type $_i$ up until month $_t$.	$\sum_{t=0}^{t_{max}} TD_{it}$	Self-Created
Installed Base	Control	MT_{it}	The number of unique machines of machine	$\sum_{t}^{t+1} MT_{it}$	PMA

			type $_i$ in operation in month $_t$.		
Unique Customers	Control	UC_{it}	The number of unique customers making use of machine type $_i$ in month $_t$.	$\sum_{t}^{t+1} UC_{it}$	PMA & CSCA
Average unscheduled downtime per machine type	Dependent	USD_{it}	The average sum of unscheduled downtime in minutes of machine type $_i$ in month $_t$	$\frac{\sum_t^{t+1} USD_{it}}{\sum_t^{t+1} MT_{it}}$	PMA
Average downtime per machine type	Dependent	DT_{it}	The average sum of unscheduled downtime and scheduled downtime in minutes of machine type $_i$ in month $_t$	$\frac{\sum_t^{t+1} USD_{it} + SD_{it}}{\sum_t^{t+1} MT_{it}}$	PMA
Mean cost of service per machine type	Dependent	COS_{it}	The sum of booked costs of machine type $_i$ in month $_t$.	$\frac{\sum_t^{t+1} COS_{it}}{\sum_t^{t+1} MT_{it}}$	CSCA & PMA

3.2.1.1 SERVICE DIAGNOSTICS TOOL VARIABLE OPERATIONALIZATION

The operationalization of the variables used to test the hypotheses in light of the SDT are shown in Table 7. The operationalizations of the sub-ordinate variables are presented in this table as well, because the values of the two dependent variables mean-time-to-use ($MTTU_{it}$) and the quality of the SDT knowledge base (QKB_{it}) are calculated using the values of these subordinate variables.

TABLE 7: THE OPERATIONALIZATION OF THE SDT VARIABLES

Variable Description	*Sort of Variable*	*Variable Symbol*	*Measured as*	*Variable Operationalization*	*Used Database*
The SDT knowledge base	Independent	CKB_{it}	The cumulative number of unique problems and fingerprints stored on machine type $_i$ up until month $_t$. Measured in absolute numbers.	$\sum_{t=0}^{t_{max}} CKB_{i,t}$	SDT

The cumulative number of PCSs	subordinate	$CPCS_{it}$	The cumulative number of unique problem fingerprint combinations that lead to a PCS for machine type $_i$ up until month $_t$. Measured in absolute numbers.	$\sum_{t=0}^{t_{max}} CPCS_{i,t}$	SDT
The cumulative number of workarounds	subordinate	CWA_{it}	The cumulative number of unique problem fingerprint combinations that lead to a WA for machine type $_i$ up until month $_t$. Measured in absolute numbers.	$\sum_{t=0}^{t_{max}} CWA_{i,t}$	SDT
The quality of the SDT knowledge base	Independent	QKB_{it}	The fraction of problem fingerprint combinations that lead to a PCS compared to all fingerprint combinations for machine type $_i$ up until month $_t$. Measured in percentages.	$\frac{\sum_{t=0}^{t_{max}} CPCS_{i,t}}{\sum_{t=0}^{t_{max}} CPCS_{i,t} + \sum_{t=0}^{t_{max}} CWA_{i,t}}$	SDT
The number of SDT usages	Subordinate	URQ_{it}	The sum of the number of first-clicks on the first given solution for a problem on machine type $_i$ in month $_t$. Measured in absolute numbers.	$\sum_{t}^{t+1} URQ_{i,t}$	UGE
The time to use the SDT	Subordinate	TTU	The difference in time between the measured time of occurrence (TOP) and measured the time of the first click (TOC) on the first proposed solution in the SDT for problem $_p$ on machine type $_i$. Measured in minutes.	$TOC_{p,i} - TOP_{p,i}$	UGE
The mean time to use the SDT	Independent	$MTTU_{it}$	The average time to use the SDT from the time of problem occurrences for all instances, where $TTU \leq 360$, for machine type $_i$ in month $_t$. Measured in minutes.	$\frac{\sum_{t}^{t+1} TTU_{i,t}}{\sum_{t}^{t+1} URQ_{i,t}}$	UGE

3.2.1.2 WORK INSTRUCTIONS VARIABLE OPERATIONALIZATION

Similarly to the SDT, the variable operationalizations for the WIs are shown in Table 8.

TABLE 8: THE OPERATIONALIZATION OF THE WI VARIABLES

Variable Description	*Sort of Variable*	*Variable Symbol*	*Measured as*	*Variable Operationalization*	*Used Database*
The WI knowledge base	Independent	CKB_{it}	The cumulative first-time usages of unique WIs on machine type $_i$ up until month $_t$. Measured in absolute numbers.	$\sum_{t=0}^{t_{max}} CKB_{i,t}$	CSCA, NC & PCD
The cumulative amount of placed comments	Subordinate	CPC_{pt}	The cumulative amount of comments placed on WI $_p$ up until month $_t$. Measured in absolute numbers.	$\sum_{t=0}^{t_{max}} CCP_{p,t}$	CMDS & CMDS.un
The cumulative amount of solved comments	Subordinate	CSC_{pt}	The cumulative amount of comments solved on WI $_p$ up until month $_t$. Measured in absolute numbers.	$\sum_{t=0}^{t_{max}} CSC_{p,t}$	CMDS
The sum of placed comments of used WIs	Subordinate	$UWPC_{it}$	The sum of the cumulative sum of placed comments up until month $_t$ for used WIs $_{pu}$ on machine type $_i$ in month $_t$. Measured in absolute numbers.	$\sum_{t}^{t+1}(\sum_{t=0}^{t_{max}} CPC_{pu,i,t})$	All
The sum of placed comments of used WIs	Subordinate	$UWSC_{it}$	The sum of the cumulative sum of solved comments up until month $_t$ for used WIs $_{pu}$ on machine type $_i$ in month $_t$. Measured in absolute numbers.	$\sum_{t}^{t+1}(\sum_{t=0}^{t_{max}} CSC_{pu,i,t})$	All
The quality level of used WIs	Independent	QUW_{it}	The fraction of the sum of the cumulative amount of solved comments and the sum of the cumulative amount of placed comments for WIs used on machine type $_I$ in month $_t$. Measured in percentages.	$\frac{UWSC_{i,t}}{UWPC_{i,t}}$	All
The number of performed actions	Subordinate	PA_{it}	The number of used materials with cost indicator > 0 on machine type $_i$ in month $_t$. Measured in absolute numbers.	$\sum_{t}^{t+1} PA_{i,t}$	CSCA
The number of used WIs	Subordinate	UW_{it}	The number of used materials with cost indicator > 0 on machine type $_I$ for which a version number existed in month $_t$. Measured in absolute numbers.	$\sum_{t}^{t+1} UW_{it}$	CSCA, NC & PCD

The completeness level	Independent	$KBCL_{it}$	The fraction of performed actions for which a WI was available on machine type $_i$ in month $_t$. Measured in percentages.	$\frac{\sum_t^{t+1} UW_{i,t}}{\sum_t^{t+1} PA_{i,t}}$	CSCA, NC & PCD
The number of used WIs validated by a TA	Subordinate	$UWTA_{it}$	The number of WIs used $_{pu}$ in month $_t$ for machine type $_i$ that are written by a TA. Measured in absolute numbers	$\sum_t^{t+1} UWTA_{pu,i,t}$	CSCA, NC, PCD & TA
The validation level	Independent	VUW_{it}	The fraction of used WIs that are validated by a TA to the total number of used WIs on machine type $_i$ in month $_t$. Measured in percentages.	$\frac{\sum_t^{t+1} UWTA_{pu,i,t}}{\sum_t^{t+1} UW_{i,t}}$	All

3.2.2 THE ECONOMETRIC MODELS

The dependent and independent variables of Table 6, Table 7 and Table 8 can be used to fill the elements of Equation 3. This is done and shown for the SDT in Equation 5 & Equation 6. The same is done for the WIs in Equation 7 and Equation 8. A graphical representation of the non-log transformed knowledge bases are shown in Figure 9 for the SDT and Figure 10 for the WIs.

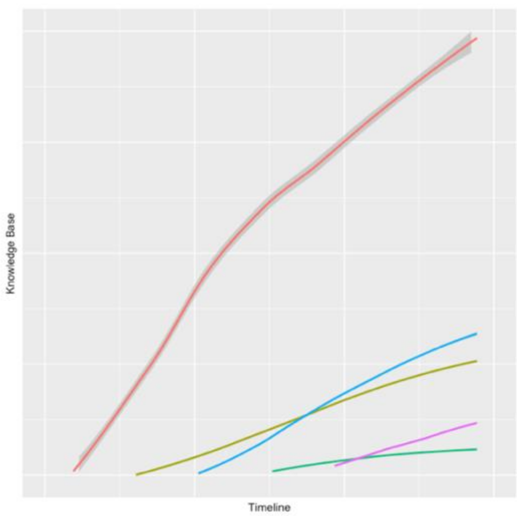

FIGURE 9: DEVELOPMENT OF THE SDT KNOWLEDGE BASE

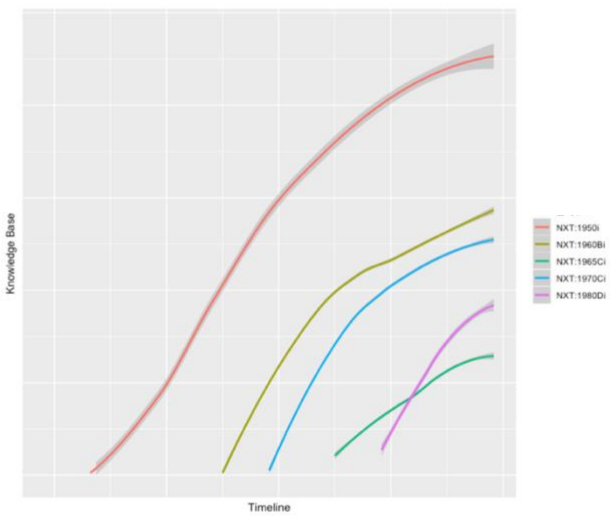

FIGURE 10: DEVELOPMENT OF THE WI KNOWLEDGE BASE

$$\ln(USD)_{i,t} = \beta_{io} + \beta_{i1} * \ln(CKB)_{i,(t-1)} + \beta_{i2} * MTTU_{i,(t-1)} + \beta_{i3} * QKB_{i,(t-1)} + \beta_{i4} * \ln(CKB)_{i,(t-1)} * MTTU_{i,(t-1)} + \beta_{i5} * \ln(CKB)_{i,(t-1)} * QKB_{i,(t-1)} + CV + a_i + u_{it}$$

EQUATION 5: MODEL FOR PREDICTING THE AVERAGE UNSCHEDULED DOWNTIME FOR THE SDT

$$\ln(COS)_{i,t} = \beta_{io} + \beta_{i1} * \ln\left(CKB_{i,(t-1)}\right) + \beta_{i2} * MTTU_{i,(t-1)} + \beta_{i3} * QKB_{i,(t-1)} + \beta_{i4} * \ln(CKB)_{i,(t-1)} * MTTU_{i,(t-1)} + \beta_{i5} * \ln(CKB)_{i,(t-1)} * QKB_{i,(t-1)} + CV + a_i + u_{it}$$

EQUATION 6: MODEL FOR PREDICTING THE MEAN COSTS FOR THE SDT

- $\beta_{io} = constant$
- $\beta_{i1} = learning\ rate$
- $\beta_{i2},\ \beta_{i3} = the\ direct\ effects\ on\ the\ dependent\ variables$
- $\beta_{i4},\ \beta_{i5} = the\ moderating\ effects\ on\ the\ dependent\ variables$

$$\ln(TD)_{i,t} = \beta_{io} + \beta_{i1} * \ln(CKB)_{i,(t-1)} + \beta_{i2} * QUW_{i,(t-1)} + \beta_{i3} * KBCL_{i,(t-1)} + \beta_{i4} * VUW_{i,(t-1)} + \beta_{i5} * \ln(CKB)_{i,(t-1)} * QUW_{i,(t-1)} + \beta_{i6} * \ln(CKB)_{i,(t-1)} * KBCL_{i,(t-1)} + \beta_{i7} * \ln(CKB)_{i,(t-1)} * VUW_{i,(t-1)} + CV + a_i + u_{it}$$

EQUATION 7: MODEL FOR PREDICTING THE AVERAGE TOTAL DOWNTIME FOR THE WORK INSTRUCTIONS

$$\ln(COS)_{i,t} = \beta_{io} + \beta_{i1} * \ln(CKB)_{i,(t-1)} + \beta_{i2} * QUW_{i,(t-1)} + \beta_{i3} * KBCL_{i,(t-1)} + \beta_{i4} * VUW_{i,(t-1)} + \beta_{i5} * \ln(CKB)_{i,(t-1)} * QUW_{i,(t-1)} + \beta_{i6} * \ln(CKB)_{i,(t-1)} * KBCL_{i,(t-1)} + \beta_{i7} * \ln(CKB)_{i,(t-1)} * VUW_{i,(t-1)} + CV + a_i + u_{it}$$

EQUATION 8: MODEL FOR PREDICTING THE MEAN COSTS FOR THE WORK INSTRUCTIONS

- $\beta_{io} = constant$
- $\beta_{i1} = learning\ rate$
- $\beta_{i2},\ \beta_{i3},\ \beta_{i4} = the\ direct\ effects\ on\ the\ dependent\ variables$
- $\beta_{i5},\ \beta_{i6},\ \beta_{i7} = the\ moderating\ effects\ on\ the\ dependent\ variables$

TABLE 9: DIFFERENCES BETWEEN PANEL DATA MODELS

	Pooled OLS	**Fixed Effects**	**Random Effects**
Functional Form	$y_{it} = \alpha + \beta * X'_{it} + v_{it}$	$y_{it} = (\alpha + u_i) + \beta * X'_{it} + v_{it}$	$y_{it} = \alpha + \beta * X'_{it} + (u_i + v_{it})$
Assumption	Individual effects, either cross-sectional or time specific, do not exist ($u_i = 0$)	Individual effects are correlated with regressors	Individual effects are not correlated with regressors
Intercepts	Constant	Varying across entities and/or time	Constant
Error Variances	Constant	Constant	Randomly distributed across entities and/or time
Hypothesis test	-	F-test	Breusch-Pagan LM test

Panel data has multiple observations of the same groups, units or entities at several different time intervals. Panel data may have individual effects, time effects or both (Park, 2011). Modeling panel data enables deeper explorations than cross-sectional or time-series data (Kennedy, 1998). As mentioned above, the used dataset is indexed on monthly measurements of the different machines that belong to the TWINSCAN platform. This means that the used panel dataset contains 5 fixed entities (*n*) with monthly observations (*T*) that make a total number of 273 monthly observations (*nT*). Furthermore, the dataset is unbalanced, which means that the dataset contains '*NAs*'. Listwise deletion of data points will be used to deal with the unbalanced dataset. Generally, there are three ways to model panel datasets (1) the pooled ordinary least squares (OLS) regression model, (2) the fixed effects model and (3) the random effects model (Park, 2011). The differences between these are shown in Table 9, which is based on Park (2011).

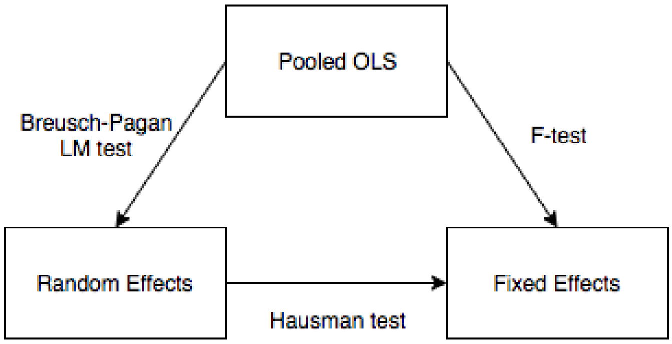

FIGURE 11: STRUCTURE FOR MODEL DECISION

In Figure 11, which is based on Park (2011), a graphical representation is given of tests that can be performed to choose the appropriate type of model. The decision to do accordingly is stepwise, one that will be discussed in the following sections. The required tests for the pooled OLS model will be performed first, because the starting point for the stepwise decision is the pooled OLS model. Thereafter, the F-test and Breusch-Pagan LM test will be performed to test whether a pooled OLS model, fixed effects model or random effects model is more appropriate. Finally, a decision between a fixed effects and random effects model must be made, which is done using the Hausman test.

3.2.3.1 VARIABLE LINEARITY

The first assumption that should be adhered to, is the assumption of linearity of the variables. This assumption states that the dependent variable is a linear function of dependent variables and the error term. Hence, the dependent variables and main independent variables should depict a linear trend. Looking at the main dependent variables USD_{it} , TD_{it} and COS_{it} , shows that these are not. Neither are the independent variables CKB_{it}. To improve estimation, these variables are converted to linear variables through performing a log-transformation, as is proposed for learning-curves by Argote (2013).

3.2.3.2 WEAKLY DEPENDENT TIME SERIES

Some sort of stability is required to understand the relationship between two variables in a linear regression that includes a time component. This means that the relationship between two variables does not change arbitrarily over time and time intervals must be stationary (Wooldridge, 2015). Furthermore, it is required that these stationary time intervals are weakly dependent upon each other. A stationary time series is weakly dependent if the correlation between x_t and x_{t+h} moves quickly enough to zero as h increases to infinity (Wooldridge, 2015). Meeting the requirement for weakly dependent time series is important to meet the requirement of the law-of-large-numbers (LLN) and the central limit theorem (CLT). The LLN is important because it states that the average from a random sample converges in probability to the population average, whereas the CLT states that the average from a random sample for any population, with finite variance when standardized, has an asymptotic normal distribution (Wooldridge, 2015). The most common situation for time series where these two criteria are not met is the situation where a variable shows a unit root, which means that they are highly persistent (Wooldridge, 2015). Testing against a unit root can be done with the augmented Dickey-Fuller test (Wooldridge, 2015). This is done and shown for all dependent variables in Table 10. All calculated p values are < 0.10, which means H_o is rejected for all dependent variables of interest.

TABLE 10: UNIT ROOT TEST RESULTS

	Augmented Dickey-Fuller Test		
Dependent variable	Unscheduled Downtime	Total Downtime	Mean Costs
H_o: Unit root			
Dickey-Fuller value	-4.0807	-4.1222	-3.2526
Lag order	12	12	12
p-value	< 0.01	0.01	0.07953

3.2.3.3 EXOGENITY

The third assumption that should be adhered to, requires the error term to have an expected value of zero in any time period. If the expected value of the error term is zero for the same time period, it is called contemporaneously exogenous. However, when the expected value is zero in any time period, it is called strictly exogenous (Wooldridge, 2015). Two probable sources of problems are omitted variables and measurement errors in the independent variables. The omitted variable problem is tried to be kept as low as possible by following the known literature concerning the subject of this research and following the expertise of ASML employees. Furthermore, the effects of the independent variables on the dependent variables should be diminished to purely the endogenous effects by adding CVs to the model. Although conceptually the omitted variable is reduced to a minimum, a problem arises concerning the empirical model of the SDT. This model, as shown in Figure 7, misses a variable concerning knowledge validation. Despite missing this variable, this research still is valuable, because the main topic of interest is the relationship between the development of a knowledge base and organizational performance.

3.2.3.4 MULTICOLLINEARITY

There should be enough variation in the samples of the independent variables, while there should not exist a linear relationship among them (Wooldridge, 2015). Testing the correlation between the two different knowledge bases, as shown in Figure 12, reveals that the increasing knowledge bases of the SDT and WIs are highly correlated (0.94, $p < 2.2e\text{-}16$). As correlations between 0.9 and 1 are considered highly correlated (Finkelstein et al., 2017), the argument to not include both variables in a single model receives support. Hence, it is preferred to have two separate empirical models for these two organizational tools and to not combine both tools into a single model.

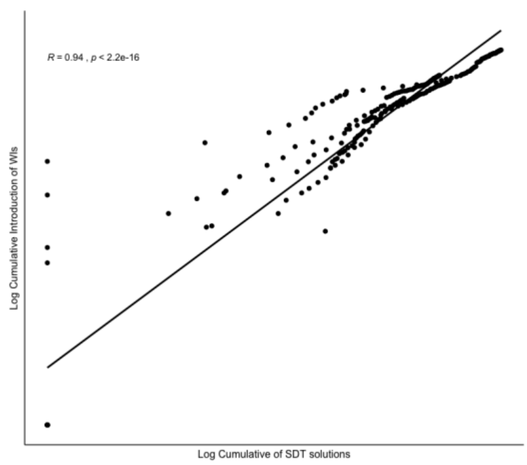

FIGURE 12: THE CORRELATION BETWEEN THE SDT AND WI KNOWLEDGE BASES

Additionally, a correlation table is created to check the dependencies among the variables of each model as presented in Equations 5 to 8. The result of this is shown in Table 11.

TABLE 11: CORRELATION TABLE OF VARIABLES

Tool		Variables						
		Knowledge Base	Control Variables			Quality Variable	Validation Variable	Utilization Variable
		Log(CKB_{it})	TD_{it}	MT_{it}	UC_{it}	QKB_{it}		$MTTU_{it}$
Service Diagnostics Tool	Log(CKB_{it})	1						
	TD_{it}	-0.01	1					
	MT_{it}	0.88	0.05	1				
	UC_{it}	0.77	-0.01	0.91	1			
	QKB_{it}	-0.53	0.01	-0.61	-0.59	1		
	$MTTU_{it}$	0.01	0.01	0.05	0	-0.18		1
		Log(CKB_{it})	TD_{it}	MT_{it}	UC_{it}	QUW_{it}	VUW_{it}	$KBCL_{it}$
Work Instructions	Log(CKB_{it})	1						
	TD_{it}	0.04	1					
	MT_{it}	0.74	0.06	1				
	UC_{it}	0.65	0.02	0.44	1			
	QUW_{it}	0.41	0.03	0.04	0.41	1		
	VUW_{it}	-0.33	0	-0.06	-0.23	-0.33	1	
	$KBCL_{it}$	0.48	0.04	0.04	0.51	0.58	-0.36	1

As can be derived from this table, the control variables UC_{it} and MT_{it} may be considered highly correlated, as they have a correlation of 0.91 (Finkelstein et al., 2017). Because of this, a decision should be made about which variable to exclude from the SDT model. To do so, the '*variance inflation factors*' (VIF scores) are consulted. VIF scores indicate multicollinearity among the variables of a regression model. Essentially, the lower the maximum VIF score, the less multicollinearity (O'Brien, 2007). By excluding one variable and comparing the VIF scores, the decision was made to exclude the UC_{it} variable, because excluding this variable produces the lowest maximum VIF score. Practically, this is the preferred variable to exclude too, since the WoW for the SDT does not differ between different customers. The VIF scores of the initial and variable excluded models can be found in Table 12, whereas an overview of the VIF scores for the variables of the WI models can be found in Appendix I.

TABLE 12: VIF SCORES FOR THE VARIABLES OF THE SDT

Variable Description	Variable Symbol	VIF Scores		
		All Variables	Excluded: UC_{it}	Excluded: MT_{it}
Time	TD_{it}	5.478425	5.084661	4.403827
Unique Customers	UC_{it}	6.656055		3.339937
Installed Base	MT_{it}	8.767856	4.399616	
Knowledge Base	CKB_{it}	7.761383	6.760481	7.748899

Knowledge Utilization	$MTTU_{it}$	1.061482	1.058864	1.061226
Knowledge Quality	QKB_{it}	2.214725	2.198421	2.214724

3.2.3.5 HETEROSKEDASTICITY AND SERIAL CORRELATION

To create a proper pooled OLS model, the assumption that the variance of the error is constant and finite is required. This must be true for all dependent variables in the same timeframe (Wooldridge, 2015). Homoskedasticity represents the situation for which this assumption is true and can be tested for by using the Breusch-Pagan test (Wooldridge, 2015). Because Koenker (1981) adjusted the Breusch-Pagan test to one that is generally preferred due to its wider applicability (Wooldridge, 2015), this version will be used. Additionally, the assumption should be tested that each moment of measurement of the independent variables is uncorrelated with the moment in time before that period. This can be tested for with the Breusch-Godfrey test (Wooldridge, 2015). If this situation is true, then the periods of time are auto-correlated with one another. Both tests are performed on the entire models as presented in Equations 5 to 8. The results of the Breusch-Pagan and Breusch-Godfrey tests are shown in Table 13 for the two tools and their two dependent variables. Results show that all models experience heteroskedasticity and the cost models experience serial correlation. According to Woolridge (2015), a common way to deal with these problems is to use robust standard errors. When both heteroskedasticity and serial correlation are present, robust standard errors should be adjusted so that they apply for both cases, as is proposed by Driscoll & Kraay (1998).

TABLE 13: RESULTS OF THE BREUSCH-PAGAN AND BREUSCH-GODFREY TEST

		Service Diagnostics Tools		*Work Instructions*	
	Dependent Variable	Unscheduled Downtime	Costs	Downtime	Costs
Breusch-Pagan	Chi-Squared	52.088	40.543	24.503	56.473
H_0: Homoskedasticity	Degrees of Freedom	7	7	10	10
	P value	5.609e-09	9.907e-07	0.006371	1.674e-08
Breusch-Godfrey	Chi-Squared	32.467	55.015	34.955	44.19
H_0: No serial correlation	Degrees of Freedom	25	25	25	25
	P value	0.1449	0.00049	0.08901	0.01033

3.2.3.6 FIXED EFFECTS MODEL

A pooled OLS model does not consider the possibility of differences between groups or time, i.e. heterogeneity across groups or time. A fixed effects model acknowledges heterogeneity across groups or time (Park, 2011). Recalling Equation 4, it shows that in a fixed effects model the unobserved effect is allowed to influence the results. To calculate this either a first-differences model, time-demeaned model or group-demeaned model can be used to show the unobserved effect (Park, 2011). The three methods are means to achieve the same result; a fixed effects model. However, the group-demeaned model, which can also be called the '*between'* model, is rarely preferred over a random effects model (Wooldridge, 2015). Choosing between a first differences model and a time-demeaned model is difficult for reasons that are not explained, as they are beyond the scope of this research. However, the time-demeaned model is used more often for cases where there is no serial correlation (Wooldridge, 2015). Since this is the case for the main dependent variables of interest, the downtime variables, it is decided to use the time-demeaned model. The time-demeaned model is also called the '*within*' method.

TABLE 14: F-TEST RESULTS

		Service Diagnostics Tools		*Work Instructions*	
	Dependent Variable	Unscheduled Downtime	Costs	Downtime	Costs
F-test	F-value	29.477	4.3846	6.3586	4.6141
H_0: **Pooled OLS is preferred**	Degrees of Freedom 1	4	4	4	4
	Degrees of Freedom 2	154	154	227	227
	P value	< 2.2e-16	0.002194	7.241e-05	0.001337

Table 14 shows the results of the F-test, which is the test that can be performed to decide between a fixed effects model and a pooled OLS model. For the F-test, the null hypothesis is that the goodness-of-fit of the pooled OLS model and fixed effects model is the same. If the null hypothesis is rejected, it may be assumed that the fixed effects model is better (Park, 2011).

3.2.3.7 RANDOM EFFECTS MODEL

In a fixed effects model, it is assumed that the unobserved effect, a_i , is correlated with the independent variables. However, in a random effects model it is assumed that the unobserved effect is uncorrelated with the independent variables used in the model. In addition, the other assumptions for a random effects model are the same ones that apply for the fixed effects model. Quasi-demeaned data on each variable is being used to calculate the random effects model instead of time-demeaned data (Wooldridge, 2015). A fixed effects model subtracts the time averages from the corresponding variable, whereas the random effects model subtracts a fraction of that time average. To decide between a pooled OLS model and a random effects model, the Breusch-Pagan Langrage Multiplier test can be performed (Park, 2011). This test checks whether the group and/or time specific variance components are zero, which is the null hypothesis. In case this null hypothesis is not rejected, a pooled OLS model is preferred. When the null hypothesis is rejected, then the random effects model is preferred. The result of the Breusch-Pagan LM tests are found in Table 15.

TABLE 15: RESULTS OF THE BREUSH-PAGAN LANGRAGE MULTIPLIER TEST

		Service Diagnostics Tools		*Work Instructions*	
	Dependent Variable	Unscheduled Downtime	Costs	Downtime	Costs
Breusch-Pagan LM	Chi-Squared	46.47	8.4832	8.0917	3.7852
$H_{0:}$ **Pooled OLS is preferred**	Degrees of Freedom	2	2	2	2
	P Value	8.111e-11	0.01438	0.01749	0.1507

3.2.3.8 FINAL MODEL DECISION

The results of the F-test and Breusch-Pagan LM test results, as shown in Table 14 and Table 15, only provide a definite answer on the type of model to use for the cost model of the WIs. Therefore, the Hausman test will not be performed for this model, as is suggested by Park (2011). To pick the final model for the other three regression models, the Hausman test should be conducted. Unless the Hausman test rejects the null hypothesis, the random effects estimates are preferred to be used for the final model (Wooldridge, 2015). The Hausman test checks whether the individual effects are uncorrelated with any regressor in the model, which is also the null hypothesis of this test (Park, 2011). When the null hypothesis is rejected, the fixed effects model is the preferred model to use (Wooldridge, 2015). The Hausman test results and the model decisions can be found in Table 16.

TABLE 16: RESULTS OF THE HAUSMAN TEST

		Service Diagnostics Tools		*Work Instructions*	
	Dependent Variable	Unscheduled Downtime	Costs	Downtime	Costs
Hausman test	Chi-Squared	3.4668	2.0449	5.447	N.A.
H_0: Random effects model is preferred	Degrees of Freedom	7	7	10	N.A.
	P-value	0.8387	0.9573	0.8594	N.A.
Model decision		Random	Random	Random	Fixed

4. RESULTS

Taking the shortcomings of the data and the appropriate models into consideration, results of the econometric Equations 5 to 8 can stepwise be build-up. In the first step, the control variables are regressed. Secondly, the main variable of interest, the knowledge base variable, is added to the model. Thereafter, the moderation variables are added without interaction effects. Finally, the interaction effects of the moderation variables are included. Through taking this stepwise approach, the robustness of the models can be verified. The results of the SDT are presented first and thereafter the results of the WIs are presented.

Note that the estimates of the direct effects of this last model are difficult to interpret, as they show the relative dependence. In other words, the estimates of the direct effects of the moderators are shown for the situation in which the volume of the knowledge base is equal to zero. Contrarily, the estimates of the direct effects of the knowledge base show the effects for when all direct moderating effects are equal to zero. Because of these reasons, the direct effect of the knowledge base as proposed in Hypothesis 1 is taken from the third model, when no indirect effects are included yet. For the moderation effects, as described in the other hypotheses, the fourth model will be consulted.

4.1 SERVICE DIAGNOSTICS TOOL RESULTS

Table 17 shows the regression result for the regression of both the unscheduled downtime and mean costs. As said before, Model 3 will be used to seek support for Hypothesis 1, since Model 4 is uninterpretable for the direct effects. Looking at Model 3, a significant decreasing relationship (-1.397, $p < 0.01$) can be observed regarding the size of the SDT knowledge base and the unscheduled downtime. This effect is also observed in Model 2. Contrarily, no significant relationship regarding the relationship between the knowledge base of the SDT and the mean costs is found. However, Model 3 shows a significant direct negative relationship for both unscheduled downtime (-0.136, $p < 0.1$) and costs (-0.171, $p < 0.1$) regarding the utilization.

TABLE 17: RESULTS OF THE SDT REGRESSION ANALYSES

Variable	Dependent Variable							
	Unscheduled Downtime				*Mean Costs*			
	Model 1	***Model 2***	***Model 3***	***Model 4***	***Model 1***	***Model 2***	***Model 3***	***Model 4***
Time	-0.0001 (0.0002)	0.0003 (0.0003)	0.002^{***} (0.0003)	0.001^{***} (0.0002)	-0.001^{**} (0.0002)	-0.0005^{***} (0.0002)	-0.0004 (0.0004)	-0.0004 (0.0004)
Installed Base	-0.001^{***} (0.0001)	0.0004 (0.0005)	0.0003 (0.0005)	0.0002 (0.0004)	0.0003 (0.001)	0.001 (0.001)	0.002 (0.001)	0.002 (0.001)
Knowledge Base		-0.201^{***} (0.076)	-1.397^{***} (0.140)	-3.404^{***} (1.125)		-0.053 (0.058)	-0.087 (0.256)	0.588 (1.228)
Quality Level			-1.027 (0.862)	-18.951^{*} (9.616)			2.064 (1.391)	7.543 (8.742)
Utilization Level			-0.136^{*} (0.077)	-0.167 (0.663)			-0.171^{*} (0.098)	-0.394 (0.609)
Knowledge Base * Quality				2.887^{*} (1.604)				-0.820 (1.420)
Knowledge Base * Utilization				0.008 (0.084)				0.028 (0.081)
Constant	5.350^{***} (0.433)	6.008^{***} (0.157)	14.384^{***} (1.197)	27.545^{***} (7.389)	9.323^{***} (0.508)	9.500^{***} (0.382)	7.311^{***} (1.461)	2.556 (8.140)
Observations	265	264	166	166	265	264	166	166
R^2	0.144	0.238	0.584	0.612	0.258	0.279	0.147	0.149
Adjusted R^2	0.137	0.230	0.571	0.595	0.252	0.271	0.120	0.112
F Statistic	21.697^{***} (df=2;262)	26.839^{***} (df=3;260)	44.907^{***} (df=5;160)	35.651^{***} (df=7;158)	44.905^{***} (df=2;262)	33.124^{***} (df=3;260)	5.488^{***} (df=5;160)	3.958^{***} (df=7;158)
NOTE:						$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$		

To find support for Hypothesis 2 and Hypothesis 4, Model 4 should be consulted. Comparing this model for both dependent variables shows that only for unscheduled downtime a moderation effect is observed (2.887, $p < 0.1$). The interpretation of continuous-by-continuous variables moderation effects can best be visualized for interpretation, as is shown in Figure 13. From this figure it can be drawn that at the early stages of the knowledge base development, PCSs have a stronger effect on lowering the unscheduled downtime than WAs. However, as the knowledge base grows, this effect diminishes and after the intersection point it becomes more desirable to add WAs to the SDT knowledge base to keep unscheduled downtimes as low as possible.

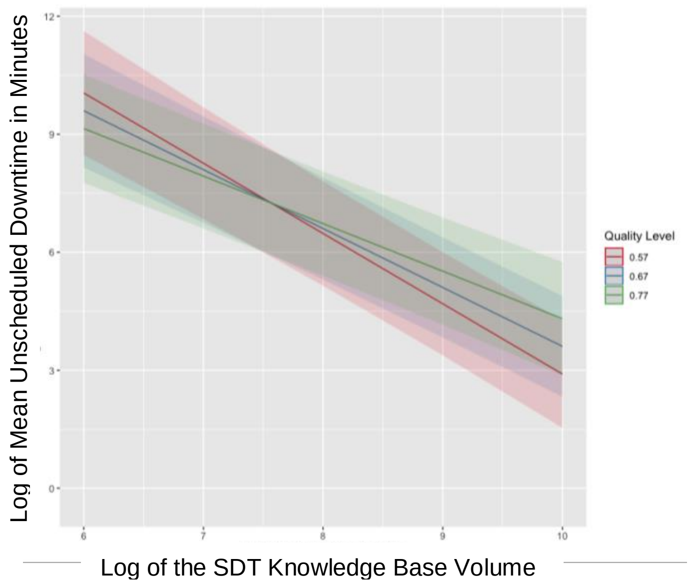

FIGURE 13: MODERATION EFFECT OF THE SDT KNOWLEDGE BASE VOLUME AND QUALITY ON UNSCHEDULED DOWNTIME

4.2 WORK INSTRUCTION RESULTS

In a similar line, conclusions for the WIs can be drawn. These are shown in Table 18. This table shows the regression results of the relationship between the WI variables and the total downtime and mean costs. Note that because the regression on mean costs was performed using a fixed effects model, the constant is not shown. From Model 3, the direct relationship between the knowledge base and these independent variables can be obtained. Doing so shows that there is a significant negative relationship between the volume of the WI knowledge base and the average total downtime in a month (-0.436, $p < 0.1$). This effect can also be observed in Model 2. Similarly to the SDT, no significant relationship was found regarding the relationship between the mean costs and the volume of the knowledge base.

Additionally, just like the SDT, a significant moderation effect between the volume of the knowledge base and the quality of the documents can be observed in Model 4. This effect exists for both total downtime (1.424, $p < 0.05$) and mean costs (1.307, $p < 0.05$). Graphical representations are created, which can be found in Figure 14 for the regression on total downtime, and in Figure 15 for the regression on mean costs. Furthermore, the results also show a significant moderation effect between the volume of the knowledge base and the validation level of the documents (-2.111, $p < 0.01$) for the regression on total downtime. A graphical representation of this effect is shown in Figure 16. Finally, no significant effects regarding the utilization of the WIs were found, as can be obtained from Table 18.

TABLE 18: RESULTS OF THE WI REGRESSION ANALYSES

	Dependent Variable							
	Unscheduled Downtime				*Mean Costs*			
Variable	***Model 1***	***Model 2***	***Model 3***	***Model 4***	***Model 1***	***Model 2***	***Model 3***	***Model 4***
Time	-0.0001 (0.0001)	0.0002 (0.0001)	0.0003^{**} (0.0001)	0.001^{***} (0.0002)	-0.001 (0.0004)	-0.0005 (0.0003)	-0.0003 (0.0004)	-0.0002 (0.0004)
Installed Base	-0.001^{***} (0.0001)	0.0004 (0.0004)	0.001^{*} (0.0004)	0.0004 (0.0003)	0.0002 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Sum of Unique Customers	-0.015^{***} (0.003)	-0.028^{***} (0.004)	-0.023^{***} (0.008)	-0.008 (0.011)	0.002 (0.023)	-0.004 (0.020)	-0.014 (0.018)	-0.012 (0.020)
Knowledge Base		-0.229^{***} (0.077)	-0.436^{*} (0.224)	0.585 (0.621)		-0.086 (0.080)	-0.187 (0.205)	-1.406 (1.022)
Quality Level			-0.069 (0.250)	-8.878^{**} (3.702)			0.170 (0.413)	-7.739^{**} (3.053)
Validation Level			0.761 (1.132)	11.265^{***} (2.702)			-0.217 (1.026)	-0.717 (5.576)
Utilization Level			-0.704 (0.453)	3.685 (2.985)			-0.009 (0.566)	0.538 (2.729)
Knowledge Base * Quality				1.424^{**} (0.603)				1.307^{**} (0.537)
Knowledge Base * Validation				-2.111^{***} (0.494)				0.038 (0.954)
Knowledge Base * Utilization				-0.853 (0.561)				-0.104 (0.553)
Constant	6.356^{***} (0.302)	7.101^{***} (0.193)	8.014^{***} (1.011)	4.418 (3.943)				
Observations	265	264	242	242	265	264	242	242
R^2	0.261	0.451	0.370	0.450	0.199	0.205	0.164	0.179
Adjusted R^2	0.253	0.442	0.352	0.426	0.177	0.180	0.124	0.128
F Statistic	30.530^{***} df=3; 261	53.071^{***} df=4; 259	19.672^{***} df=7; 234	18.890^{***} df=10; 231	21.281^{***} df=3; 257	16.430^{***} df=4; 255	6.429^{***} df=7; 230	4.945^{***} df=10; 227
NOTE:						$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$		

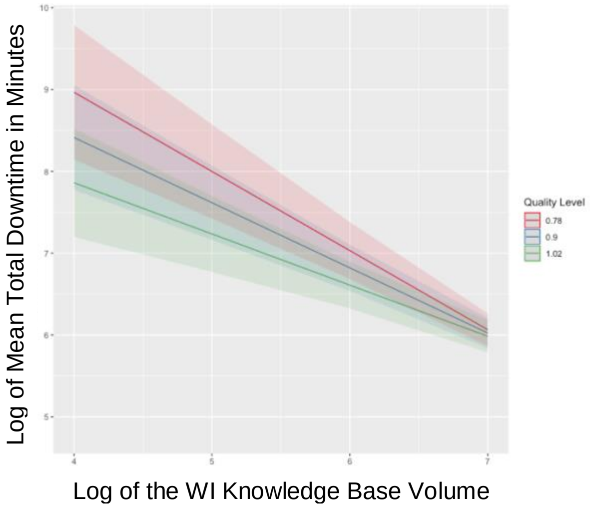

FIGURE 14: MODERATION EFFECT OF THE WI KNOWLEDGE BASE VOLUME AND QUALITY ON TOTAL DOWNTIME

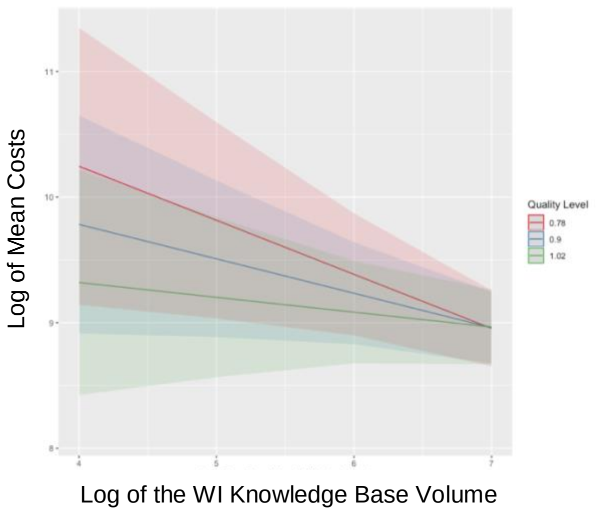

FIGURE 15: MODERATION EFFECT OF THE WI KNOWLEDGE BASE VOLUME AND QUALITY ON MEAN COSTS

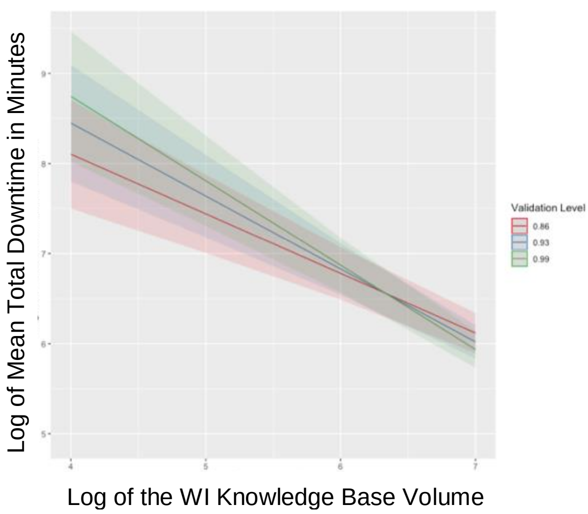

FIGURE 16: MODERATION EFFECT OF THE WI KNOWLEDGE BASE VOLUME AND VALIDATION ON MEAN TOTAL DOWNTIME

5. DISCUSSION

The found and presented results in the previous chapter can be used as input for discussion. This will first be done in relation to the consulted theories and the theoretical framework and thereafter regarding the specific ASML situation. Finally, research limitations and directions for further research will be discussed.

5.1 THEORETICAL IMPLICATIONS

The obtained results of this research show that the amount of knowledge stored in organizational IT tools significantly contributes to an improvement in organizational performance, although this relationship is only observed in relation to the reduction of downtimes and not for costs. Hence, with this relation partial support for Hypothesis 1 is found. A possible explanation of this observation may be found in the nature of the investigated tools. Both tools are intended to support the CS department with achieving their primary goal which is minimizing downtime, or alternatively optimizing machine availability. For the CS department, costs are a secondary concern. Herewith, cost reduction is expected to happen as a by-product of more effective labor and part orders that are initially intended to decrease downtimes.

Additionally, a moderation effect regarding the volume knowledge base and the quality of that knowledge base is observed for both WI regression models and for the SDT regression model on unscheduled downtime. Unlike the initially proposed hypothesis, in which performance is expected to improve when quality is higher, a different relationship is found. Performing a simple slope analysis on these three moderation effects leads to the same pattern observation. The effect of high quality on performance diminishes when the amount of knowledge in an organizational IT tool increases. Thus, high quality is not beneficial for performance at all knowledge base volumes. However, despite differently than expected, this research finds enough support to partially accept Hypothesis 2.

Furthermore, a possible moderation effect between the volume of a knowledge base and the validation level of that knowledge base could only be tested for the WIs. Despite this given fact, a moderation effect is observed for the WIs regarding the total downtime. However, unlike proposed in Hypothesis 3, this effect shows a different relationship than expected. A high validation level is not beneficial for the organizational performance at all knowledge base volume sizes, but it becomes more important as the knowledge base grows. One possible explanation for this may again be found in the maturity of the knowledge. In the early stages the WI documents are more likely to be subject to change than in the later stages. Hence, validating the content of the WIs by TAs can quickly be overturned by a major update. This thought needs further research, which should then be performed on a document level rather than the machine level that this research has followed.

Finally, no support for Hypothesis 4, which concerns the relationship between the knowledge base volume and the utilization of that knowledge base, has been observed in the empirical results. However, a direct relationship between the SDT knowledge base utilization and the unscheduled downtime was found, but this effect is not shown for the WIs. Despite this finding, Hypothesis 4 must be rejected.

5.2 MANAGERIAL IMPLICATIONS

The obtained results not only have theoretical implications, but may also guide ASML to increase their organizational performance. Firstly, advise for the SDT will be given. Thereafter, the same will be done for the WIs.

5.2.1 SDT IMPLICATIONS

The main conclusion regarding the SDT is in line with the theoretical finding that a bigger volume of the SDT knowledge base leads to smaller unscheduled downtimes. For each additional percental growth of the SDT volume, the unscheduled downtime decreases with 1.397%. Hence, to achieve an optimal performance of the SDT, the input of problem-fingerprint combinations with a solution should be kept as high as possible. Within ASML, this is represented by the share-rate. The share-rate is the fraction of analyzed-problem fingerprints for which a given solution is either validated or a new solution is given by an engineer. Encouragement policies to increase the share-rate are already in place and this research can be used to further emphasize the importance to create new solutions or link existing solutions to problem-fingerprint combinations that do not have a solution yet.

Additionally, policies concerning the share rate should promote the creation of PCSs rather than WAs at the beginning of developing a knowledge base, because of the observed quality moderation effect. Encouragements to create PCSs can gradually be let go until the intersection point has been reached. After intersection, creating WAs instead of PCSs should be encouraged within the share-rate policy. Looking back at Figure 13, this intersection occurs at around 2000 documents. This means that the production of PCSs should be promoted until 2000 problem-fingerprint combinations are stored in the SDT. After 2000 problem-fingerprint combinations, the production of WAs should be promoted. This is relatively early, since the SDT knowledge base for the NXT:1950i is around ten times as big at the last point of the used dataset. However, there are two possible explanations for this finding. Firstly, the diminishing effect of PCSs compared to WAs may be due to the capturing of the low-hanging fruits first. In other words, in the beginning of developing SDT content, relatively easy, high impact and often reoccurring problems are given solutions first. Thereafter, solutions for less well understood, lower impact and less reoccurring problems are developed. Secondly, the nature of PCSs and WAs may be of influence. PCSs often advise to replace a part, whereas WAs often advise to restart the system or certain drivers. It is assumable that replacing a part takes more time than restarting the system or drivers and as a result it may be more preferable, in terms of downtime, to have more WAs than PCSs for problems that do not require an immediate replacement of parts.

Lastly, an important finding is that the mean time to use the SDT was found to be of significant importance to reduce both unscheduled downtime and costs. This is true for all knowledge base volume sizes, since no moderation effect was found. Currently, the use of the SDT is only mandatory for all problems that take $\geq$ 2 hours to resolve, but this research finds support to change this policy into one that suggests to always first consult the SDT when a problem has occurred. To support this policy change, the argument that for every minute quicker to use the SDT since the occurrence of the problem, a reduction of unscheduled downtime by 0.136% and costs by 0.171% can be used. Naturally, the time to pick up the case by an engineer was not taken into consideration, but the results emphasizes the need to react swiftly when a customer requires the help of ASML.

5.2.2 WI IMPLICATIONS

The main finding regarding the WIs is similar to the other findings, i.e. for every additional percental increase of the volume of the WI knowledge base, a 0.436 % decrease in total downtime is expected. Based on this finding it is suggested to keep developing WIs for the NXT machines until all parts of a machine have a corresponding WI.

Additionally, the importance of the quality of the WIs diminishes as the knowledge base grows, which is similar to the findings concerning the SDT quality. However, unlike the SDT, the intersection point for the WI quality is not reached as quickly as for the SDT quality. Only after 1100 uniquely available WIs does a low quality level become preferable over a high quality level in terms of total downtime and after 1000 uniquely available WIs does this happen in terms of costs. These numbers, which can be derived from Figure 14 and Figure 15, have only just been met for the NXT:1950i. Therefore, it is practically suggested to keep an emphasis on solving comments on the WIs for all TWINSCAN machines except for the NXT:1950i. Two possible explanations for the found moderation effect regarding quality are proposed. Firstly, like the SDT, the low hanging fruits are captured first. In other words, it is assumable that solving the first comment on a WI has a higher impact on improving the quality of a WI than any other comments thereafter. Secondly, it is assumable that engineers writing the WIs get better in doing so as they get more experience writing them. Hence, solving comments on WIs that are written in later stages do not have as much impact as solving comments on WIs that are written in earlier stages of the knowledge base development.

Furthermore, the observed moderation effect regarding the validation of the WI knowledge base shows that it becomes increasingly important to have TAs validating WIs. To optimally make use of the validation mechanism, should the fraction of validated WIs gradually increase until reaching a 100% validation level at around 550 uniquely available WIs. These numbers can be obtained by looking at Figure 16. A possible explanation for this finding may be found in the number of authors that are involved in writing these WIs at different stages of the knowledge base. In the early stages of developing the knowledge base, relatively little people are involved. As a result, it may be assumable that the variability between the different WIs is low and validating them does not have as much impact as when many different authors are contributing to the WI knowledge base.

Finally, the results regarding the utilization of the WI knowledge base shows inconsistent results. Possibly this is due to the made assumption that a WI is always used to perform an action when there is a WI available. Likely will it be insightful to have actual data concerning WI utilization, since the SDT utilization shows to have a significant direct impact on organizational performance. To develop more accurate data on WI utilization within ASML, the creation of a back-end database that offline monitors the clicks performed to open a WI for the first time on a day is advised. This can be done through monitoring the clicks performed to open a WI for the first time on a day, much like how the usage of the SDT is monitored. The generated metadata can be shared with ASML servers when the content of the WIs is being updated. Although the generated utilization data is still likely to be biased, will it be a step closer to monitoring actual WI utilization.

5.3 LIMITATIONS & FURTHER RESEARCH

Although this research has contributed to creating an understanding of the investigated relationship between organizational IT tools and organizational performance in light of organizational learning theories, further questions and shortcomings have come to light while conducting this research. Firstly, although a case-study provides a rich and thorough understanding of a problem, the reliability, validity and generalizability of such a research is often questioned (Flyvbjerg, 2006). Although Flyvberg (2006) tackles this and other misunderstandings about case-studies, another method for increasing the reliability, validity, and generalizability of the results provided in this research exists. Increasing these elements can be done through conducting different case-studies to validate, reject or support the presented findings of this research. Such researches could have different starting points, of which two will be mentioned. Firstly, recalling Figure 12, the development of the knowledge base volumes of the two investigated tools is highly correlated. This is true despite the tools having different learning mechanisms that lay at the origin of the knowledge that is generated and stored in these tools. Investigating the reasons behind the high correlation of the knowledge base developments through including the different learning mechanisms may be interesting for further research and multiple questions can be used as a starting point. For example, is the observed correlation present because both tools share the same learning mechanism, learning-by-others, as origin of the knowledge generation? Or is this relationship found because the majority of the knowledge stored in these tools is generated through that learning mechanism? Secondly, questioning whether the knowledge base development of other ASML organizational IT tool databases are equally correlated, and whether they have a similar contribution to other organizational performance metrics, could be interesting. Taking this question outside the boundaries of ASML for comparison with other organizations and their organizational IT tools, can be done to validate the generalizability and increase strength of the theoretical findings, if similar results are obtained.

Furthermore, two possible conceptual improvements can be identified. Firstly, this research has only looked at the explicit knowledge and has not taken tacit knowledge into consideration. Recalling the earlier discussed continuum between tacit and explicit knowledge, the effects of knowledge in organizational IT tools on organizational performance are likely to be affected by the tacit knowledge of an engineer. For example, it may be assumable that more experienced engineers are more likely to rely on their own knowledge, instead of using the investigated tools. Including engineer knowledge in future research on the relation between organizational knowledge in IT tools and organizational performance is likely to produce valuable additional insights. Secondly, although not significant, the results on costs for the knowledge base volume still hint towards an effect of the volume of a knowledge base on costs. There may be other factors, which were not taken into account, that influence the relationship between the development of the knowledge base and the performance metric booked costs.

Finally, four possible improvements concerning data selection and handling must be discussed. Firstly, the SDT empirical model is missing data concerning validation of the knowledge base. Although this does not appear to be a problem for the interpretation of the main effect of interest, it is suggested to take this data into consideration in future researches since more data will be available by then. Secondly, because each problem-fingerprint combination is taken as a single document input, a great overlap between multiple problem fingerprints may exist. Clustering problem-fingerprints and corresponding solutions, with the help of text-mining, is expected to overcome this problem in future researches. By doing so, it is also expected that the intersection of the quality moderation effect of the SDT knowledge base takes place in a relatively later stage of the knowledge base development, as each additional cluster will add more relative value. Thirdly, text-mining analysis can also be of value for the WIs to determine quality of the WIs based on their content, such as the number of words or used images. However, for such a research should the appropriate IT infrastructure be in place to extract historic versions in batch. Fourth, no distinction between the different types of comments, typographical, technical and configurational that can be placed on WIs, was made. Hence, all three types of comments are expected to equally impact quality of a WI, but it may be assumable that a typographical comment has less impact on the quality than a configurational comment. Taking this into consideration in further research is expected to deliver more accurate results.

6. CONCLUSION

This research has investigated the effects of knowledge, which is obtained through organizational learning processes, stored in organizational IT tools on the organizational performance. It has started by posing the question: "*how does a growing knowledge base in organizational IT tools contribute to an improved organizational performance?*" The results of this research show that the growing number of documents, the volume of the knowledge base, contributes to an increased organizational performance. Additionally, factors that possibly influence this relationship were investigated by posing the question: "*what factors influence the relationship between knowledge stored in organizational IT tools and organizational performance?*" The results of this research show that this effect is influenced by the quality and validation level of the knowledge base. Furthermore, hints pointing towards the importance of organizational IT tools utilization were obtained, but these hints should be investigated more thoroughly in future research.

REFERENCES

Alavi, M., & Leidner, D. E. (2001). Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *Management Information Systems Quarterly*, *25*(1), 107–136. http://doi.org/10.2307/3250961

Alsina, E. F., Chica, M., Trawiński, K., & Regattieri, A. (2018). On the use of machine learning methods to predict component reliability from data-driven industrial case studies. *International Journal of Advanced Manufacturing Technology*, *94*(5–8), 2419–2433. http://doi.org/10.1007/s00170-017-1039-x

Argote, L. (2011). Organizational learning research: Past, present and future. *Management Learning*, *42*(4), 439–446. http://doi.org/10.1177/1350507611408217

Argote, L. (2013). *Organizational Learning: Creating, Retaining and Transferring Knowledge* (Second Edi). New York: Springer.

Argote, L., & Epple, D. (1990). Learning Curves in Manufacturing. *Science*, *247*(4945), 920–924. http://doi.org/10.1126/science.247.4945.920

Argote, L., Mcevily, B., & Reagans, R. (2003). Framework and review of emerging themes managing knowledge in organizations: An integrative framework and review of emerging themes. *Management Science*, *49*(August 2014), 571–582. http://doi.org/10.1287/mnsc.49.4.571.14424

Argote, L., & Miron-Spektor, E. (2011). Organizational Learning: From Experience to Knowledge. *Organization Science*, *22*(5), 1123–1137. http://doi.org/10.1287/orsc.1100.0621

Baines, T., & Lightfoot, H. W. (2013). Servitization of the manufacturing firm: Exploring the operations practices and technologies that deliver advanced services. *International Journal of Operations & Production Management*, *34*(1), 2–35. Retrieved from https://doi.org/10.1108/IJOPM-02-2012-0086

Brown, J. S., & Duguid, P. (2001). Knowledge and Organization. *Organization Science*, *12*(2), 198–213. http://doi.org/10.1287/orsc.12.2.198.10116

De Liso, N., Filatrella, G., & Weaver, N. (2001). On endogenous growth and increasing returns: modeling learning-by-doing and the division of labor. *Journal of Economic Behavior and Organization*, *46*(1), 39–55. http://doi.org/10.1016/S0167-2681(01)00186-X

Devaraj, S., & Kohli, R. (2003). Performance Impacts of Information Technology : Is Actual Performance Impacts of Information Technology : Is Actual Usage the Missing Link ? *Management Science*, *49*(March 2015).

Dorroh, J., Gulledge, T., & Womer, N. (1994). Investment in knowledge: A generalization of learning by experience. *Management Science*, *40*(8), 947–958. http://doi.org/10.1287/mnsc.40.8.947

Driscoll, J. C., & Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *The Review of Economics and Statistics*, *80*(4), 549–560.

Fadel, K. J., Durcikova, A., & Cha, H. S. (2009). Information Influence in Mediated Knowledge Transfer. *International Journal of Knowledge Management*, *5*(4), 26–42. http://doi.org/10.4018/jkm.2009062902

Finkelstein, A., Harman, M., Jia, Y., Martin, W., Sarro, F., & Zhang, Y. (2017). Investigating the relationship between price, rating, and popularity in the Blackberry World App Store. *Information and Software Technology*, *87*, 119–139. http://doi.org/10.1016/j.infsof.2017.03.002

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, *12*(2), 219–245. Retrieved from https://www.researchgate.net/profile/Bent_Flyvbjerg/publication/286140682_Five_misunderstandings_about_case-study_research_in_Qualitative_Research_Practice/links/56902bf008aed0aed810f3ed.pdf

Gold, A. H., Malthora, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems*, *18*(1), 185–214. http://doi.org/10.1002/ceat.201000522

Greene, W. H. (2003). *Econometric analysis* (5th Eition, Vol. 5). Prentice Hall. http://doi.org/10.1198/jasa.2002.s458

Haas, M. R., & Hansen, M. T. (2005). When using knowledge can hurt performance: The value of organizational capabilities in a management consulting company. *Strategic Management Journal*, *26*(1), 1–24. http://doi.org/10.1002/smj.429

Kane, G. C., & Alavi, M. (2007). Information Technology and Organizational Learning: An Investigation of Exploration and Exploitation Processes. *Organization Science*, *18*(5), 796–812. http://doi.org/10.1287/orsc.1070.0286

Kennedy, P. (1998). *A Guide to Econometrics* (4th Editio). MIT Press. http://doi.org/10.15713/ins.mmj.3

Kimmerle, J., Cress, U., & Held, C. (2010). The interplay between individual and collective knowledge: technologies for organisational learning and knowledge building. *Knowledge Management Research & Practice*, *8*(1), 33–44. http://doi.org/10.1057/kmrp.2009.36

Koenker, R. (1981). A Note on Studentizing a Test for Heteroscedasticity. *Journal of Econometrics*, *17*(1), 107–112. http://doi.org/10.1016/0304-4076(81)90062-2

Lapré, M. A. (2010). Inside the Organizational Learning Curve: Understanding the Organizational Learning Process. *Technology, Information and Operations Management*, *4*(1), 1–103. http://doi.org/10.1561/0200000023

Lex. (2015). ASML: a serious market leader. Retrieved September 27, 2018, from https://www.ft.com/content/6f90926a-bc2e-11e4-b6ec-00144feab7de

Marwick, A. D. (2001). Knowledge management technology. *IBM Systems Journal*, *40*(4), 814–830. http://doi.org/10.1147/sj.404.0814

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of Information and System Quality: An Empirical Examination Within the Context of Data Warehousing. *Journal of Management Information Systems*, *21*(4), 199–235. http://doi.org/10.1080/07421222.2005.11045823

Nonaka, I., & von Krogh, G. (2009). Perspective—Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *Organization Science*, *20*(3), 635–652. http://doi.org/10.1287/orsc.1080.0412

O'Brien, R. M. (2007). A Caution regarding Rules of Thumb for Variance Inflation Factors. *Quality and Quantity*, *41*(5), 673–690. http://doi.org/10.1007/s11135-006-9018-6

Park, H. M. (2011). *Practical Guides To Panel Data Modeling: A Step by Step Approach*. *Public Management and Public Analysis Program*. Retrieved from http://www.iuj.ac.jp/faculty/kucc625/documents/panel_iuj.pdf

Poston, R. S., & Speier, C. (2005). Effective Use of Knowledge Management Systems: A Process Model of Content Ratings and Credibility Indicators. *Management Information Systems Quarterly*, *29*(2), 221–244.

Ryu, C., Kim, Y. T., Chaudhury, A., & Rao, H. R. (2005). Knowledge Acquisition via Three Learning Processes in Enterprise Information Portals: Learning-by-Investment, Learning-by-Doing, and Learning-from-Others. *Management Information Systems Quarterly*, *29*(2), 245–278. http://doi.org/10.2307/25148679

Schaller, R. R. (1997). Moore's Law: past, present, future. *IEEE Spectrum*, *34*(6), 52–59. http://doi.org/10.1109/6.591665

Schoeffler, S., Buzzell, R. D., & Heany, D. F. (1974). The impact of strategic planning on profit performance. *Harvard Business Review*, 137–145. http://doi.org/10.1016/0024-6301(74)90244-1

SEMI. (2004). *Standard for Definition and Measurement of Equipment Reliability , Availability , and Maintainability (Ram)*. San Jose, California.

Teece, D. J. (1998). Capturing Value from Knowledge Assets: The New Economy, Markets for Know-how, and Intangible Assets. *California Management Review*, *40*(3), 55–79.

Tippins, M. J., & Sohi, R. S. (2003). IT competency and firm performance: Is organizational learning a missing link? *Strategic Management Journal*, *24*(8), 745–761. http://doi.org/10.1002/smj.337

Tsoukas, H. (1996). The firm as a distributed knowledge system: A constructionist approach. *Strategic Management Journal*, *17*(Winter Special), 11–25. http://doi.org/10.1002/smj.4250171104

Valtakoski, A. (2017). Explaining servitization failure and deservitization: A knowledge-based perspective. *Industrial Marketing Management*, *60*, 138–150.

Vandermerwe, S., & Rada, J. (1988). Servitization of Business: Adding Value by Adding Services. *European Management Journal*, *6*(4), 314–325.

Wang, Y., Huang, Q., Davison, R. M., & Yang, F. (2018). Effect of transactive memory systems on team performance mediated by knowledge transfer. *International Journal of Information Management*, *41*(April), 65–79. http://doi.org/https://doi.org/10.1016/j.ijinfomgt.2018.04.001

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach* (Sixth Edit). Boston: Cencage Learning. Retrieved from http://econoxpert.com/wp-content/uploads/2017/01/Wooldridge-Introductory-Econometrics_-A-Modern-Approach-6th-Edition-c2016.pdf?i=1

APPENDIX I

TABLE 19: VIF SCORES OF THE WI VARIABLES

Variable Name	Variable Symbol	VIF score
Time	TD_{it}	4.495665
Unique Customers	UC_{it}	2.011610
Installed Base	MT_{it}	4.540266
Knowledge Base	CKB_{it}	4.568865
Knowledge Quality	QUW_{it}	1.285597
Knowledge Validation	VUW_{it}	1.723318
Knowledge Completeness	$KBCL_{it}$	2.218703