UNIVERSITETET I AGDER

Towards an understanding of big data analytics as a weapon for competitive performance

Frank Danielsen

Vetle Augustin Framnes

SUPERVISORS

Dag Håkon Olsen

Polyxeni Vasilakopoulou

Patrick Mikalef

University of Agder, 2017
Faculty of Social Sciences
Department of Information systems

"You can have data without information,
but you cannot have information
without data."

Daniel Keys Moran

Foreword

This master’s thesis report is the results of the work performed by two students from the master's degree programme in information systems at the University of Agder. Our collaboration started earlier when we worked together on our bachelor’s degree project and we continued working together throughout our master’s programme coursework.

The *big data* topic was something that caught our attention early in the first semester of the master's programme and our interest in the field has only grown, so much so that we decided to work on *big data* related themes whenever possible. Therefore, *big data* became our natural choice when choosing a theme for our master’s programme thesis.

The possibilities, the hype and the interest, from both an industry and academic perspective piqued our curiosity, which makes this topic an exciting choice. We are glad to work on relevant applied research that is needed and can be used to further understand the world of *big data*.

The progress we have made, and our ability to direct this research, is only possible because of the well-constructed master’s programme and the education provided by the supportive instructors and the edifying coursework. We especially want to thank our supervisors, Professor Dag Håkon Olsen (UiA), Associate Professor Polyxeni Vasilakopoulou (UiA) and Postdoc Patrick Mikalef (NTNU) who provided great help, guidance and encouragement. They were also great sparring partners throughout the project.

Finally, we want to thank our families and loved ones. Thanks for letting us devote time to this thesis even at the cost of precious family time. All academic articles that were reviewed, endless academic conversations about *big data* and countless hours of study have been a sacrifice you have also had to endure. It is clear that we would not be where we are without your loving support.

Kristiansand 29. May 2017

Frank Danielsen Vetle Augustin Framnes

Frank Danielsen Vetle Augustin Framnes

Abstract

Context: *Big data* has in the recent years been an area of interest among innovative organizations and has started to become a major priority for organizations in general, either through their own *big data* departments or by purchasing *big data* analyses from suppliers. *Big data analytics* means more knowledge from more data sources and is by many prophesied to be a contributing source of big change in how organizations receive their intelligence.

Purpose: This thesis investigates the connection between *big data* and *competitive performance*. This connection could be explained through the following two paths; 1) how *big data analytics* contribute to making an organization more agile/dynamic and 2) how *big data analytics* improves daily operations. To measure this, we looked at *big data analytics capabilities*, *dynamic capabilities* and *operational capabilities* in addition to *competitive performance*.

Methods: The methods that were used in this research was mainly of a quantitative type in addition to a qualitative case study and a two-phased literature review. We had to establish how to define and measure *big data analytics capabilities*. To do this we had to collect and review existing literature on *big data analytical capabilities*. We then had to do the same process with *dynamic capabilities, operational capabilities* and *competitive performance*. Even if there were little to no examples of previous literature on the whole scope of our research area, there were jigsaw bits that contained important knowledge on the different parts of the research area. With the help of previous literature and our case study, a survey was created. This was sent to big organizations in Nordic countries, mainly from the Kapital 500 list of the biggest organizations in Norway and Forbes Global 2000 list, where we focused on the biggest organizations in the Nordic countries. Extensive work was put into sorting away organizations that did not use *big data,* and to get respondents that did. A total of 135 respondents completed the survey and 107 of those used *big data* solutions. We developed a model with four hypotheses to investigate the relationship between *big data* analytic capabilities and *competitive performance* through the mediating concept of *dynamic capabilities* and *operational capabilities*. We analysed the responses using structural equation modelling. Specifically, we used partial least square path modelling (PLS-SEM). The tool used to distribute the survey was SurveyGizmo and we used SmartPLS to analyse the data.

Results: Our analyses validated our first hypothesis which points to the positive correlation between *big data analytics capabilities* and *dynamic capabilities* to be significant. Further, the second hypothesis that stipulates the path from *dynamic capabilities* to *competitive performance* was significant. We also found significance on our third hypothesis which suggest a positive correlation between *big data analytics capabilities* and *operational capabilities*. We failed to find any significance on the fourth hypothesis which proposed that there is a positive correlation between *operational capabilities* and *competitive performance*. Also, we did not

find any significance on environmental factors moderating effect on the second and fourth hypothesis.

Conclusion:

Overall, our results shows that the concept of *big data analytics capability* is transformed into *competitive performance* through the path of *dynamic capability* which can be seen as a mediating factor. This study contributes to better understand how *big data analytics* investments are turned into competitive actions and will be particularly valuable for companies using *big data*.

Key words: big data analytics; big data analytics capabilities; dynamic capabilities; operational capabilities; competitive performance.

Table of contents

List of illustrations

List of tables

1. Introduction

During the first Techonomy Conference in 2010, Eric Schmidt, Google's CEO at the time, said: "there were 5 exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days." (Kirkpatrick, 2010). In other words, we generate volumes of personal and public data at a rapidly expanding rate. Due to worldwide diffusion, use of mobile devices, social networks and the advancement of "Internet of Things", we leave behind an astonishingly large digital footprint every day. This abundance of public data has attracted interest amongst scholars and practitioners who refer to the phenomenon as *big data* (Wamba, Akter, Edwards, Chopin & Gnanzou, 2015).

The phase *big data* is more than just a catchphrase that describes a large amount of data. Today, data is generated at the speed of thought and via a wide variety of sources that collects and disseminates the structured, semi-structured and unstructured information stream (Russom, 2011). These kinds of characteristics pose a major challenge for companies who wants to extract insight from such disordered public data. Technological advances, and particularly *big data analytics*, have nevertheless made it possible to retrieve information from every type of data. *Big data analytics* can create business value through increased transparency, predicting customer needs, creating adaptive business models and supporting, or even replacing, human decision making (Wamba et al., 2015). *Big data analytics* can thus provide organizations with a competitive edge through the creation of an information-based arsenal.

There is a clear association between a company's performance and their competitive advantage. Research has shown that IS investments alone cannot improve a company's performance metric, but established IT capabilities may at define a company's performance (Kim, Shin, Kim & Lee, 2011). We are not alone in our belief that companies need to develop distinctive *big data analytics capabilities* in order to capitalize on their *big data analytics* investments (Gupta & George, 2016; Wamba et al., 2017).

Furthermore, we take the stance adopted by other scholars who believe *dynamic capabilities* and *operational capabilities* can explain the source of sustained competitive advantage.

The purpose of this study is to understand the mechanisms that unite *big data* and *competitive performance* by proposing the following research question:

"By what paths are big data analytics capabilities transformed into Competitive Performance"

The research question was answered through an extensive study consisting of two phases. The first phase was initiated to establish a conceptual framework and model that could later be tested for generalisability to a larger population. To do this, we studied theory through a systematic literature review and collected empirical data through six semi-structured depth interviews with companies that had adopted *big data*.

In the second phase, data from 107 *big data*-using companies in the Nordic region were collected through a survey and then analysed.

1.1 Key concepts

'Big data' and 'Big data analytics' are two terms often used by academia and business organizations. We use the following definitions for the remainder of the thesis:

Big data is about gathering, storing, managing and accessing the necessary data for analytics (Espinosa & Armour, 2016). The data could consist of large structured, semi structured and unstructured data sets that require new forms (untraditional solutions) of processing capability to enable better decision making (Emani, Cullot & Nicolle, 2015; Garmaki, Boughzala & Wamba, 2016; Wamba et al., 2015). Examples include e.g., social media, radio-frequency identification (RFID) tags, smart phones, and sensors (Gupta & George, 2016).

Big data analytics is about technologies (e.g., database and data mining tools) and techniques (e.g., analytical methods) that a company can employ to analyse large scale, complex data (Kwon, Lee & Shin, 2014) and report insights not attainable with past data technologies (Garmaki et al., 2016). These technologies can include data management (massively parallel-processing databases), often open-source programming (Hadoop, MapReduce), statistical analysis (sentiment analysis, time-series analysis), advanced visualization tools that help structure and connect data to uncover hidden patterns, anomalies, unknown correlations, and other actionable insights.

1.2 Limitations

This research is focused on how *big data analytics capabilities* affect *dynamic* and *operational capabilities* and indirectly affects *competitive performance*. We chose those two paths based on previous literature. It may most likely exist other paths that *big data analytic capabilities* correlates with *competitive performance* but we chose to examine dynamic and operational sides of an organization.

We measure competitive performance by asking respondents to self-report on questions related performance in comparison to their competitors. There might be other ways to measure performance, for instance, to view the profitability from each organization or by viewing other facts. The problem that occurs by doing it this way might be "old numbers", unavailable data or wrongfully information. Then again, self-reported answers might contain biased answers. Ultimately, the literature support the self-reporting as the best way to measure our variables.

Our survey contains quite a few questions, a total of 67. Our model is quite massive and to secure enough respondents we had to minimize measurements on our variables, often resulting in three questions per variable (with a few that only had two questions).

1.3 Motivation and benefits

Organizations know that they can attain great value from *big data*. The first step in achieving this goal is to develop knowledge about the capabilities an organization could acquire versus what results to expect. In addition to satisfying our personal curiosity, we also enhance our *big data* competencies. There is also great satisfaction in knowing that we are contributing to *big data* research, especially when it is a needed, wanted and still a relatively unexplored area.

1.4 Contributions

Since this master thesis focuses on the association between *big data analytics capabilities* and *competitive performance*, the expected output of this work is to better understand how organizations use *big data* and how much this increases their *competitive performance*. Also, to identify the kind of factors that affect or moderate the links will broaden the understanding and strengthen the results. We believe this is a somewhat uncharted territory and a study is needed to obtain a deeper understanding of *big data* solutions in organizations.

1.5 Content and structure

The rest of this report is constructed as follows; Chapter two constitutes the theoretical foundation that this research relies on. In chapter three we propose our conceptual model along with the hypotheses. Chapter four summarizes the applied methodology used throughout the study. Chapter five presents our findings and in chapter six, these findings are reflected upon by using prior literature practical implications. Lastly, chapter seven contains the conclusions of this study. The report then presents the references and attachments.

2. Theoretical foundation

The theoretical foundation of this thesis is based on an extended literature review and an empirical, explorative case study on the use of *big data analytics* in Norwegian and Italian companies. The first part of the literature review (part A) was conducted in parallel with the case study and the second part (part B) was carried out subsequently.

Part A answered the following research question: *"what assets facilitate the use of* big data *solutions"*? This is a fundamental question to understand the factors that lead to *big data analytics* initiatives. The same question was also used for the exploratory case study that formed the basis for the article "Big data analytics capabilities: Antecedents and Business Value" (Mikalef, Framnes, Danielsen, Krogstie & Olsen, 2017) which were submitted and accepted into the Pacific-Asia Conference on Information Systems 2017 conference (PACIS 2017, 2017).

Reflections made during this early phase led to further refinement of the research question, which guided us to expand the theoretical point of view. Part B was then carried out to further define *competitive performance*, *dynamic capabilities* and *operational capabilities*. We became aware of these concepts during part A.

The literature review and the exploratory case study were conducted through a rigorous and systematic process. The results from these studies were used to develop the overarching research model. These research model components are detailed in the sub sections below and they include: *big data analytics capability*, *dynamic capability*, *operational capability*, *competitive performance* and *environmental factors*.

2.1 Big data analytics capability

To acquire business value from *big data analytics* investments requires investments into *big data analytics capabilities* (Gupta & George, 2016). Garmaki et al. (2016) define BDAC as an organization's ability to exploit the combination of data and IT components with the goal of achieving competitiveness. In recent years, several researchers have focused their attention on systematically sorting and understanding these capabilities. In part A of the literature review we found broad support for Wamba et al. (2017)'s suggested model of *big data analytics capabilities* The model included three principle factors: expertise capabilities, management capabilities, technology skills. As we reviewed additional theories, we discovered new capabilities that were important. These include organizational skills (e.g., data-driven culture, organizational agility or organizational analytics capabilities), presentation capabilities and top management support. These competencies did not fit into Wamba's framework. Therefore, we decided to use Gupta and George (2016) 's framework as the basis for our description of

BDAC. We define the elements of Gupta and George (2016) framework in the sections that follow.

2.1.1 Tangibles

Gupta and George (2016) describe tangible resources as items “that can be sold or bought in a market”. The paper mentions examples like financial resources and physical assets. These assets are then divided into basic resources, data and technology.

Data

An important part of *big data analytics capabilities* is access to *big data*. Business organizations are becoming more and more aware of the value of data. Gupta and George (2016) surmise that the growth in business data utilization rate is directly related to the *big data* growth rate. Five data types are identified: public data, private data, data exhaust, community data, and self-quantification data.

The definitions of these data types are source-specific. For example, Gupta and George (2016) refer to accessible data (e.g., non-personal data) when they define the following data types.

- *Public data* are often free data provided by either governmental institutions, private organizations or individuals.
- *Private data* are organization-owned data.
- *Data exhaust* represents data with no or little value in its own context but might provide valuable intel when connected to other data.
- *Community data* are, for instance, Facebook, Twitter and other social media generated data.
- *Self-quantification data* are data generated from wearable technologies like smart watches, fitness bands and the like (George, Haas & Pentland, 2014).

Data can be further divided into external and internal data:

- *Internal data* are organizational data created by the organizational processes. Examples are inventory updates, sales, transactions or other internal processes.
- *External data* are data from external sources, either public, private but achievable through buying or trading, community data among others (Gupta & George, 2016; Zhao, Fan & Hu, 2014)

Gupta and George (2016, p. 4) state that “firms interested in creating *big data analytics capabilities* must integrate their internal and external data”. Often, business do not want to share proprietary data, even within its own organisation. Departmental issues can also bring about resistance to sharing and merging data across the organization. These problems highlight the need for support from business leaders to establish norms and standards that facilitate more transparent data policies. Particularly in cases when data will not only benefit individual departments, but become available to the whole company (GalbRaith, 2014). *Big data analytics* solutions are performing *now-casting analytics*, which is a prediction of the present. In order to accurately now-cast, data access should be real-time or as close to real-time as possible (Akter, Wamba, Gunasekaran, Dubey & Childe, 2016).

Technology

In addition to have access to data from all above-mentioned sources, there is a need for technology that is able to handle and support analytical processes of those data, even if it is gigantic, diverse and fast-moving (Gupta & George, 2016). The systems should also support data handling that is required to, for instance, handle data as a data lake, where organizations collect data at a fast pace, both from internal and external sources.

Gupta and George (2016) State that as much as 80% of the data that companies hold have an unstructured format. As a result, relational database technology becomes an insufficient solution (Garmaki et al., 2016) More sophisticated technological solutions are needed in this case. Technologies such as Hadoop, which is a Java-based open-source framework, consequently appeared to tackle this type of data (Gupta & George, 2016). This is because many *big data* technologies support Hadoop. Emani et al. (2015) suggest using technologies that easily interface with Hadoop. This is also supported by interviews from our exploratory case study. Scalable infrastructure options are needed because of the data growth; a high data growth rate often requires parallel extensions of computer power (Emani et al., 2015).

Basic resources

Gupta and George (2016) explain the importance of investing both time and economical resources into *big data analytics capabilities*. Interviewees from our exploratory case study confirmed this and explained that since *big data* is a relative new phenomenon, the nature of how you operate analytical processes with *big data* requires testing, learning, searching, failing and patience. Tallon, Ramirez and Short (2013) further suggest that investments into *big data* should be separate from other IT investments since these *big data* investments, if bundled with other IT investments, might undermine other strategic IT projects.

2.1.2 Human resources

Technical skills (technical knowledge) and managerial skills (technology management capability, business knowledge and relational knowledge) are the two most important skills as stated by Gupta and George (2016) and other prior IT capability researchers (Kim et al., 2011; Wamba et al., 2017). Further description of these two groups of skills follow.

Technology skills

This refers to skills required to extract value or intelligence from *big data*. Competencies like 1) understanding, operating and even modifying systems like Hadoop, 2) machine learning, 3) data scrubbing, 4) statistical analysis and 5) data extraction are examples of these skills (Gupta & George, 2016). Those abilities are rare for individuals trained in non-technical business fields and difficult to absorb, both for an organization and for individuals who are not trained in science, technology, engineering and mathematics (STEM). This is therefore a major business challenge for organizations who want to benefit from *big data* since they have to either hire new employees with those competencies and train them to learn about the non-technical field or invest in an existing employee to learn the required competencies by sending them to school (Chen, Chiang & Storey, 2012). Although this challenge might change or disappear over time, acquiring these advanced competencies will require significant investment, patience and time. Journeyman-level technical skills that are more commonplace, like competencies in database management, typesetting languages, soft skills analysis or business enterprise systems are also skillsets that fall under *big data* management (Akter et al., 2016). These types of skills are easy to extend in a short period by hiring recent business school graduates with specific technical coursework or training current employees via online classes or short courses/seminars (Gupta & George, 2016).

Managerial skills

Unlike most technology skills, managerial skills are developed on-the-job and over time. Managerial skills are also industry-specific, tacit and heterogeneously dispersed across organizations (Gupta & George, 2016). This is an important component of *big data analytics capabilities* since managers need to have a good understanding of how to apply *big data analytics*. results. There needs to be good communication and collaboration between the management team, management levels and department staff to fully exploit the business value *big data analytics* creates. Akter et al. (2016) suggest that *big data* management solutions are necessary to support business-oriented goals. They suggest that there are two types of managers: technology managers and "ordinary" managers. Both managers need to have good business understanding and knowledge. Wamba et al. (2015) point out that managers need to ensure that data scientists are familiar with and concerned about typical business topics/issues. They should also know how to use appropriate business vocabulary and have good interdepartmental communication skills to better understand how to use, track and report *big data analytics* results (Janssen, van der Voort & Wahyudi, 2017).

2.1.3 Intangibles

Teece (2015, p. 119) suggests that intangible resources are central to an organization's performance, especially in dynamic markets. These resources have no clear or visible boundaries and are highly context-dependent (Mata, Fuerst & Barney, 1995; Teece, 2014). They do not "travel" well and are hard to mimic. Exceptions do exist though, like tradable copyrights, trademarks, patents and the like. Gupta and George (2016) divide intangible resources into two sub categories: organizational learning and data-driven culture. This will be explained in the following paragraphs.

Organizational learning

One side of organizational learning is the ability to reconfigure resources according to changes brought on by possibilities or forced by external dynamic environmental situations. *Big data* solutions might predict market trends that encourage internal changes to achieve continuous *competitive performance.* Therefore, the added value drives the organizational need for (Grant, 1996; Teece, Pisano & Shuen, 1997). Gupta and George (2016) argue that data analytics does not tell the whole story and organizations with higher organizational knowledge will likely have an advantage when making decisions based on or supported by *big data analytics* results. Another view of organizational learning involves improving an organization's day to day processes and learning how to incorporate *big data analytics* into those processes. The more organizations can see and perfect the seamless integration of *big data*, the higher competitive advantage they might achieve. This was discovered in our exploratory case study while interviewing several organizations in the media industry. Media organizations that had an established system in place to use *big data* solutions in their daily operations had an advantage over the ones that were lagging. There was a broad agreement that, in this industry, it is necessary to use and make progress with *big data analytics* usage. In fact, their survival depends on it.

Data-driven culture

Organizational culture is an intangible and is very hard to understand and describe; therefore, it is difficult to replicate. The definition among some researchers proclaim that organizational culture is the glue of an organization, while others say it encompasses almost all areas of an organization (Gupta & George, 2016). Recent research supports that organizational culture is tightly connected with success of *big data* initiatives (LaValle, Lesser, Shockley, Hopkins & Kruschwitz, 2011; Ross, Beath & Quaadgras, 2013). To get the most value from *big data*, there must be an inherent trust in the results: that they are correct, and that intuition and personal beliefs don't eclipse or contradict the decision-making process that is based on analytical results; particularly when decisions are influenced by individuals, as there might not be any personal gain from *big data* investments (Gupta & George, 2016; McAfee, Brynjolfsson, Davenport, Patil & Barton, 2012; Quaadgras, Ross & Beath, 2013).

2.2 Dynamic Capability

A key question in the field of strategic management is how organizations can achieve and sustain competitive advantage (Teece et al., 1997). This question has led to many different theories. One of the theories currently receiving a lot of attention is the 'resources-based view of the firm' (RBV), which sees firm resources as the source of competitive advantage. By obtaining resources that are valuable, imperfectly mobile and heterogeneously distributed across firms, the firm can attain superior performance and a sustained competitive advantage (Mata et al., 1995). However, RBV has received criticism, over the years, for ignoring factors that surround these resources. Essentially, the opinion is that RBV provides information that is static and does not provide solutions or give guidance about how future valuable resources should be acquired or how the resource base can be renewed. This shortcoming has fuelled the theoretical footing of the *dynamic capabilities* concept. The concept emerged in the 1990's and has since progressively evolved to become one of the most influential approaches to management research of our time (Ambrosini & Bowman, 2009; Schilke, 2014b). The *dynamic capabilities* theory aims to explain how organizations can continually acquire valuable, competitive resources that match or change the marketplace (Wheeler, 2002).

The concept has, since it was first introduced, been developed further. The stream of literature has shared several definitions and concept breakdowns. Teece et al. (1997, p. 517) define *dynamic capabilities* as *"the ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments"*. Winter (2003, p. 991) define *dynamic capabilities* as *"those that operate to extend, modify or create ordinary capabilities"*. Another, more recent definition supported in the literature is based on the work by Helfat et al. (2009, p. 1), which states: *"the capacity of an organization to purposefully create, extend or modify its resource base"* (Fainshmidt, Pezeshkan, Frazier, Nair & Markowski, 2016; Kim et al., 2011; Protogerou, Caloghirou & Lioukas, 2012). The definitions show that there is a consensus about the DC construct. This reflects that organizational processes are on a common basis, and have the role of changing the company's resource base.

In a time of global markets, new technologies arise and new competition is bred. Modern companies must be alert and able to respond to potential threats or opportunities in the market by developing and using *dynamic capabilities* (Roberts, Campbell & Vijayasarathy, 2016). New and innovative solutions may put an end to established ways of doing work and businesses that are unable to adapt, could be eradicated. By using *dynamic capabilities*, businesses can become adaptable and escape unfavourable path dependencies, thereby achieving evolutionary fitness (Teece, 2007). The more often companies engage in sensing and transforming activities, the better their *dynamic capabilities* will become and the more it will be integrated into the organizational memory (Wilden & Gudergan, 2015). In environments characterized by change, even *dynamic capabilities* can become worthless. Companies therefore need to evaluate and renew their capabilities from time to time (Schilke, 2014b) .

There has also been an ongoing discussion about what environment *dynamic capabilities* yield most value from. Teece et al. (1997) pointed out that there is an obvious value in obtaining *dynamic capabilities* in rapidly changing environments. Several researchers have extended this point-of-view by expressing that *dynamic capabilities* are primarily of value in turbulent environments (Ferrier, Holsapple & Sabherwal, 2010; Wilden & Gudergan, 2015). According to Pavlou and El Sawy (2011), the literature has presumed that *dynamic capabilities* have no value in stable environment. They may even be detrimental. They further point out the misconception by showing that *dynamic capabilities* can only have a positive impact throughout the whole spectrum of environmental turbulence.

The concept is often divided into three connected activities, which are sensing, seizing and transforming.

Sensing

An important part of *dynamic capabilities* is the ability to sense threats and opportunities in the environment. In order to discover such threats or opportunities, it is essential for companies to frequently scan both 'local' and 'distant' markets and technologies (Roberts et al., 2016; Teece, 2007). To identify opportunities, employees must have access to information about the entire business ecosystem and be able to understand latent demands from customers. The information used as a basis may originate from a variety of sources, such as a conversation at an industry meeting, from news or feedback from frustrated customers (Teece, 2007). The use of technology can provide a lot of valuable information and thus strengthen the organization's sensing ability (Roberts et al., 2016).

Seizing

The sensing capability is of no value if the organization is unable to respond to what is being observed. An equally important characteristic of *dynamic capabilities* is therefore the ability to seize the identified opportunities. Seizing activity implies evaluation of various options to accommodate the identified opportunity. The option should acquire marketplace acceptance (Wilden, Gudergan, Nielsen & Lings, 2013). Careful planning is an essential part of seizing and it usually involves more intensive resources than the sensing activity, which can be a low-cost activity (Teece et al., 1997).

Transforming

To benefit from sensing and seizing activities, the company's asset orchestration must be reconfigured to achieve better utilization. The transforming characteristic of *dynamic capabilities* implies the company's ability to adjust their capabilities in response to changes (Wilden & Gudergan, 2015). Organizational change is usually a costly affair, so companies must develop effective change processes to mitigate low pay-off changes. The ability for a company to transform and reconfigure its resource base is itself a learned skill. The more experience and the more practice, the more manageable it becomes (Teece et al., 1997; Wilden & Gudergan, 2015).

2.3 Operational Capability

An organization's *operational capabilities*, sometimes referred as ordinary capabilities, is a collective description of an organization's ability to "make a living" (Drnevich & Kriauciunas, 2011) or to convert inputs into outputs (Wilden & Gudergan, 2015). Contrary to *dynamic capabilities*, *operational capabilities* is connected to technical fitness and not evolutionary fitness (Li, Shang & Slaughter, 2010). Higher *operational capabilities* helps organizations to execute operations more efficiently and therefore achieve greater technical fitness (Li et al., 2010). Many organizations do not understand that IT can enable *dynamic capabilities*. The focus on IT is as an enabler of high-level *operational capabilities*. *Operational capabilities* only offers short-term temporary advantages and organizations could probably lose those advantages as changes occur in the environment (El Sawy & Pavlou, 2008). Teece (2007) found that if an organization lacked *dynamic capabilities*, it could still make a (good) competitive return for a <u>short</u> period. It cannot sustain "supra-competitive" returns in the long-term. Winter (2003, p. 992) has said "[the] archetypical enterprise [have] competencies/resources but [lack] *dynamic capabilities* [that] will in equilibrium 'earn a living by producing and selling the same product, on the same scale and to the same customer population'". Then again, if there is a significant, tacit, non-inimitable component of an organization's superior operational competence, it has the potential to support superior performance, even for a limited time (Teece, 2007). Wu, Melnyk and Flynn (2010) support this by saying that *operational capabilities* has emerged gradually over time and is not easy to mimic. Often, it is transferred to future generations by teaching (i.e. internships). Wilden and Gudergan (2015) divides *operational capabilities* into marketing and technological capabilities. Marketing capabilities is, for instance, market knowledge, customer relationships and distribution channels. Technological capabilities consider the efficiency of the internal processes in an organization, how well they handle day to day operations, technical expertise and equipment. When improving *operational capabilities*, the industry 'best practise' may provide greater results and advantages, even though "best practises" are widely imitable, and don't offer heterogeneity (Drnevich & Kriauciunas, 2011). *Operational capabilities* appears to have a foundation of a firm's operations (Drnevich & Kriauciunas, 2011).

2.4 Competitive performance

The abovementioned capabilities (*big data analytics capabilities*, *dynamic capabilities* and *operational capabilities*) affect an organization's performance. Research literature often focus on performance, to some degree, as our literature review revealed. Several different ways were proposed to define performance and measure it. Some methods define and measure performance (e.g. financial performance, operational performance, market performance or product/service performance) as increased/decreased efficiency and productivity based on history (Chang & Gurbaxani, 2013; Pavlou & El Sawy, 2011). Other methods proclaim that the best way to define performance is via comparison to relevant competitors (Ferrier et al., 2010; Rai & Tang, 2010). Other terms and definitions have surfaced, for instance, *competitive performance* (Ferrier et al., 2010; Lim, Stratopoulos & Wirjanto, 2011; Lu & Ramamurthy, 2010; Rai & Tang, 2010) or *competitive advantage* (Drnevich & Kriauciunas, 2011; Li & Liu, 2014; Li et al., 2010). A description of *competitive performance* provided by Rai and Tang (2010) states that it is an ability to capture market share, remain profitable, keep growing, and be innovative and cost-efficient in comparison to major competitors. Sometimes the time dimension is also accounted for in, for example, the overall financial performance over the past few years (Kim et al., 2011). Organizations might have a competitive advantage in some business activities and disadvantages in others. For instance, they could have advantages when it comes to achieving product effectiveness (quality and innovativeness) and disadvantages when it comes to process efficiency (time to market and low cost) (Drnevich & Kriauciunas, 2011). Therefore, these factors should always be considered since *competitive performance* is a complex concept.

2.5 Environment factors

Studies have supported the notion that the effect of capabilities on performance can be affected or moderated by environmental factors (Chen, Preston & Swink, 2015; Chen et al., 2014; Rai & Tang, 2010; Wilden et al., 2013). Both in RBV theory and in *dynamic capabilities* theory, scholars have conducted empirical investigations on the effects environmental factors have on organizations (Chen et al., 2014; Li & Liu, 2014). Some scholars divide the organizational environment into dynamism (or stability), complexity (or simplicity) and hostility (or munificence) (Chen et al., 2014; Li & Liu, 2014). Environmental factors can also moderate *big data analytics capabilities* (Chen et al., 2015), *operational capabilities* (Pavlou & El Sawy, 2011; Wilden & Gudergan, 2015) and *dynamic capabilities*. While some scholars proclaim that environmental factors decide if *dynamic capabilities* have organizational value (Chen et al., 2015; Wilden et al., 2013), others have provided research that shows *dynamic capabilities* can add value through the entire spectrum of environmental turbulence (Li & Liu, 2014). Pavlou and El Sawy (2011) have also done research to establish the positive moderation effect that environmental factors have on the relationship between *dynamic capabilities* and *competitive performance*.

Environmental hostility

Organizations operate in environments with some grade of hostility. Zahra and Garvis (2000) describes this observation as the existence of unfavourable external forces in the organization's business environment. This can again lead to, for instance, radical changes in the industry, more intense regulatory burdens or fierce rivalry among competitors (Chen et al., 2014; Dess & Beard, 1984; Werner, Brouthers & Brouthers, 1996). Other impacts could be high taxes, lack of knowledge and education in the population, which leads to a lack of staff competence, fragile infrastructure, economic instability and workforce insecurity. Chen et al. (2014) argues that this might hinder developing capabilities (e.g., IT capabilities) which in turn staggers an organization's ability to be agile and flexible. Chen further explains that hostile environments also lead to greater restrictions in communication, formal procedures and centralization of strategic decision-making. This could have a negative effect on both *big data analytics capabilities*, where *big data* might be prevented in lightening decision-making and easing strict centralization of decision-making, and *dynamic capabilities*, where agility might stagger because of rules, procedures and slow, formal decision making.

Environmental dynamism

An environment that is dynamic might be a negative moderator between an organization's ordinary, day to day processes and *competitive performance*. Li and Liu (2014, p. 2795) explains that "dynamism is interpreted as unpredictability, that is, the rate of change and innovation in an industry as well as the uncertainty or unpredictability of actions by customers". Other researchers describe *environmental dynamism* as the rate of unpredictability of an environment. Dynamism can mean changes to, for instance, product/service obsolescence, technology change, competitors' moves, and shifts in customer demand (Chen et al., 2014). In *dynamic capabilities* theory, *environmental dynamism* is a key situational parameter that affects the grade of correlation between *dynamic capabilities* and *competitive performance* (Chen et al., 2015). Teece et al. (1997) propose that *dynamic capabilities* are directly counteracting environmental change, which makes creating a competitive advantage difficult since many changes occur simultaneously. Maintaining previously gained competitive benefits (Chen et al., 2014). Li and Liu (2014) also supports this statement by adding that in very high competitive environments, resources are difficult to obtain and therefore the agility of an organization can contribute to short term advantages. On the other hand, less competitive environments might not hinder long term advantages even in an organization that lack agility. So, even if there might be some use of *dynamic capabilities* in a less turbulent environment, scholars, like Li and Liu (2014), point to the natural connection between *environmental dynamism* and *dynamic capabilities*.

Environmental complexity

Environmental complexity is the heterogeneity and diversity of external factors. It occurs in terms of diversity of customer buying habits, nature of competition and product lines (Chen et al., 2014; Newkirk & Lederer, 2006). Competition might enable or strengthen *environmental complexity* since organizations struggle in environments that is comprised of finite resources. Thus, the more competitors there are, the higher the competitive intensity among the organizations (Wilden et al., 2013). Managers of organizations with complex environments are concerned with more factors than those that operate in more simple environments (Chen et al., 2014). Thus, managers have more difficulties with their decision making in the higher-complexity environments. Chen et al. (2014) also found that managers have difficulties making fundamental changes and opt for making smaller-scale decisions in these environments.

3 Conceptual model and hypotheses

We created a research model based on our research question, the literature review and our exploratory case study. This model represents the relationships among key constructs in our research area. To help explain the connection between the variables in the model, a hypothesis was constructed. The following sections describe the model and theory in detail.

3.1 Conceptual model

To ensure the quality of our model, we developed a model that leverages prior (empirical and theoretical) research so that all the elements of our model were extensions of established principles. Early in the project, we designed a trial model that we could adapt as our knowledge expanded. By developing the model in this adaptive way, we ensured that our research model did not disregard prior work in the field. Additionally, implementing our own empirical work and data increases the reliability and validity of our model. See figure 1 for the conceptual model.

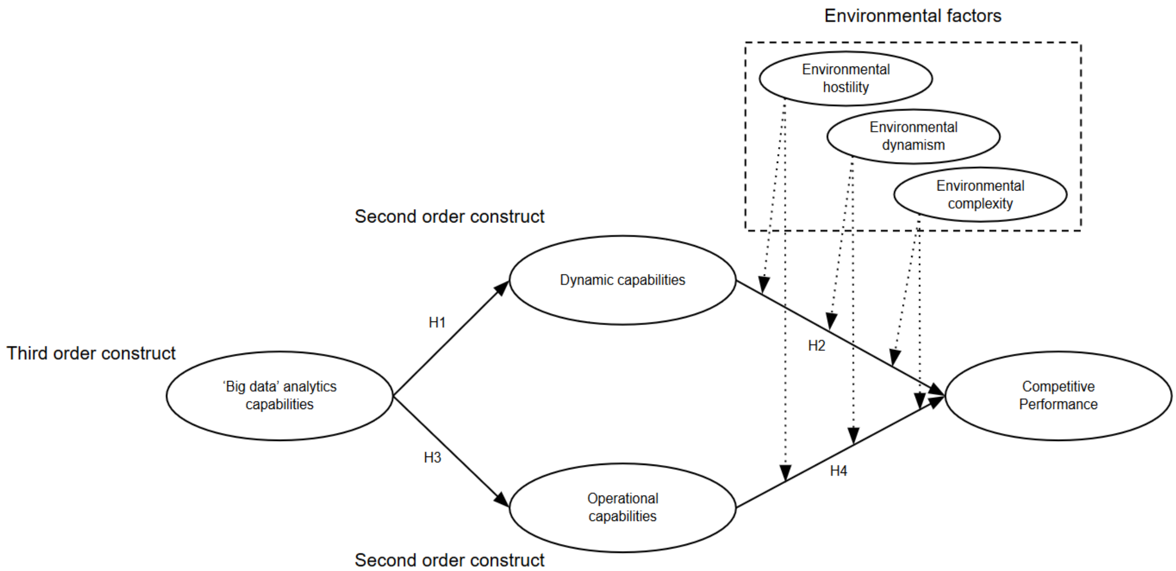


Figure 1: Conceptual model

3.2 Hypotheses

The next step after developing a model is to determine a hypothesis. The connection between *big data* (or *big data analytics capabilities*) and performance (or *competitive performance*) could now be explained by mediating variables. To be able to test if there was a correlation effect between the model's elements, four hypotheses were created. These are presented in the following section.

Hypothesis 1

In today's technological business environment, we leave digital footprints in almost every activity we perform. If analysed, these remaining bits and bytes may provide great business insight. However, the vast amounts of information surpass the human processing ability. By using information technology to tackle such informational resources, firms can obtain knowledge assets (Chi, Ravichandran & Andrevski, 2010). Assets like these enhance the sensing, seizing and transforming capabilities of managers; enabling them to better capture market intelligence, accelerating their decision-making processes, making better decisions based on real-time data, respond to changes in the market and swiftly reconfiguring the resource base accordingly (Chi et al., 2010; Roberts et al., 2016).

By using *big data analytics*, companies are capable of extracting insight from vast amounts of recurring data and a variety of formats (Garmaki et al., 2016).

Success stories of *big data analytics* implementations in the literature has created a gold-rush-like atmosphere (Espinosa & Armour, 2016). Some companies' adoption of *big data analytics* can be explained by the concept of isomorphism, which denotes the pursuit of similarities with competing companies. Moreover, some companies' IT innovation and adoption is affected by the adoption of the same concept in a well-known company, regardless of the rationality of doing so (Kwon et al., 2014). Business insight does not emerge automatically simply by applying technical *big data analytics* solution to data (Sharma, Mithas & Kankanhalli, 2014). To capitalize on *big data* and *big data analytics* investments, companies should invest in *big data analytics capabilities* which incorporates the organizational ability to utilize data assets (Espinosa & Armour, 2016; Garmaki et al., 2016).

We therefore postulate the following hypothesis:

H1: "There is a positive correlation between big data analytics capabilities and dynamic capabilities"

Hypothesis 2

The potential *competitive performance* that companies gain by developing their *dynamic capabilities* does not likely stem from functionality (e.g. various forms of analytical tools). *competitive performance* is usually acquired in the open market and is thus accessible by competitors. The potential value is located in the new resource configuration that is derived from the insight that *dynamic capabilities* provides. By being able to sense, seize and transform the resource base according to the threats and opportunities in the market, companies can continue to search for new temporary advantages (Chen et al., 2015). *Dynamic capabilities* can increase the speed, effectiveness and efficiency of the company to better accommodate for upheavals in the environment (Drnevich & Kriauciunas, 2011). Outstanding *dynamic capabilities* are therefore expected to improve the firm's *competitive performance* (Kim et al., 2011).

We therefore postulate the following hypothesis:

H2: "There is a positive correlation between dynamic capabilities and competitive performance"

Hypothesis 3

In addition to using *big data analytics* to enhance an organization's ability to be agile, we believe *big data analytics* can be used to enhance the organization's day to day routines. When looking at *big data analytics* as a source of information, those analytic results can be used in, for instance, advertising placement or by sales staff needing "evidence" when generating sales (this was an example that surfaced during interviews in our explorative case study). Another example is that *big data analytics* can lead to greater optimization of transportation resources. This could therefore enable faster and better asset utilization over time (Chen et al., 2015). As Gupta and George (2016) suggest, *big data analytics capabilities* affects the results provided by *big data analytics* and as Kim et al. (2011) also suggest, capabilities derived from IT generally affect performance through *operational capabilities*.

We therefore postulate the following hypothesis:

H3: "There is a positive correlation between big data analytics capabilities and operational capabilities"

Hypothesis 4

Stronger *operational capabilities* is a more efficient and economical way of performing day to day tasks. It enables organizations to surpass less strong organizations, which provides the strong company a competitive advantage (Li et al., 2010; Roth & Jackson III, 1995; Wilden & Gudergan, 2015; Wu et al., 2010). This advantage is temporary, short term (El Sawy & Pavlou, 2008; Teece, 2007). *operational capabilities* can provide an organization with technical fitness but not evolutionary fitness (Teece, 2007). By using *big data analytics* based solutions an organization can achieve even stronger *operational capabilities* (e.g. by advertising placements, better assets resource scaling or other intelligence related to increase efficiency and decrease costs). This was observed in our exploratory case study as well as supported in the literature review (Chen et al., 2015).

We therefore postulate the following hypothesis:

H4: "There is a positive correlation between operational capabilities and competitive performance"

4. Research method

In this chapter, we explain the choice of method used to answer our research question, which is *"By what paths are big data analytics capabilities transformed into competitive performance "*.

"A method is a procedure, a means of solving problems and developing new knowledge. Any means that serves this purpose belongs to the arsenal of methods" (Hellevik, 2011, p. 12, Our own translation).

In accordance with the above definition, we first present the research approach, research design and timeline. Then we present phase one. This phase contains methods used in the initial exploratory case study, the literature review process, operationalising the model's variables and a description of how we planned to secure reliability and validity of the survey. After this, we present phase two. This phase contains methods for collecting and analysing data. Model reliability and validity is also presented here, but this time in more detail and with focus on the analysing stage. Finally, we explain our view and goals regarding research ethics.

4.1 Research approach

Initially, we performed an exploratory qualitative study to become more familiar with the concept and the phenomenon of *big data* in organizations. This study had an intensive research strategy where we examined a small group of organizations in depth. The results of this study and the parallel conducted literature review (part A) helped further shape the agenda for the research in this study.

To answer the formulated research question, we considered it appropriate to use a quantitative approach with an extensive research strategy where we examined many units with few variables (Hellevik, 2011, p. 111). Thus, a priority of width rather than depth. The study can be seen as deductive, where our hypotheses are based on theoretical knowledge from the literature. As we want to test our model that was deductively divided we adopted a quantitative approach. This approach fits the requirements for collecting empirical evidence that can be used to evaluate the hypotheses and thus enlighten research question.

4.1.1 Survey

In this study, we have mainly used a survey as a strategic approach to answer our research question. The study is also a triangulation of the strategies since we have applied both a case study and a survey strategy. The survey aims to obtain the same kind of data from a larger group of people in a systematic manner and thereby looks for statistical patterns and ultimately generalizes the results for a larger population (Oates, 2006, pp. 35, 37).

4.2 Research design

The research design shows our action plan to collect and process data needed to answer our research question. As shown in the graphical representation (see figure 2), the plan was divided into two phases in addition to a completion phase for refinement. The first phase (Phase I) consisted of the initial exploratory case study, the literature review, the conceptual model development and the collection protocol and data analysis plan. In other words, all the planning and preliminary work was accomplished prior to the data collection. The second phase (Phase II) consisted of the execution of the data collection and -analysis. As soon as the survey was sent out, it was no longer possible to make changes. We therefore worked a lot in Phase I to ensure that we would collect the appropriate data. The transition to the second phase can thus be a "point of no return". Throughout both phases, we worked on structuring and building this report so that it would reflect the work that was done.

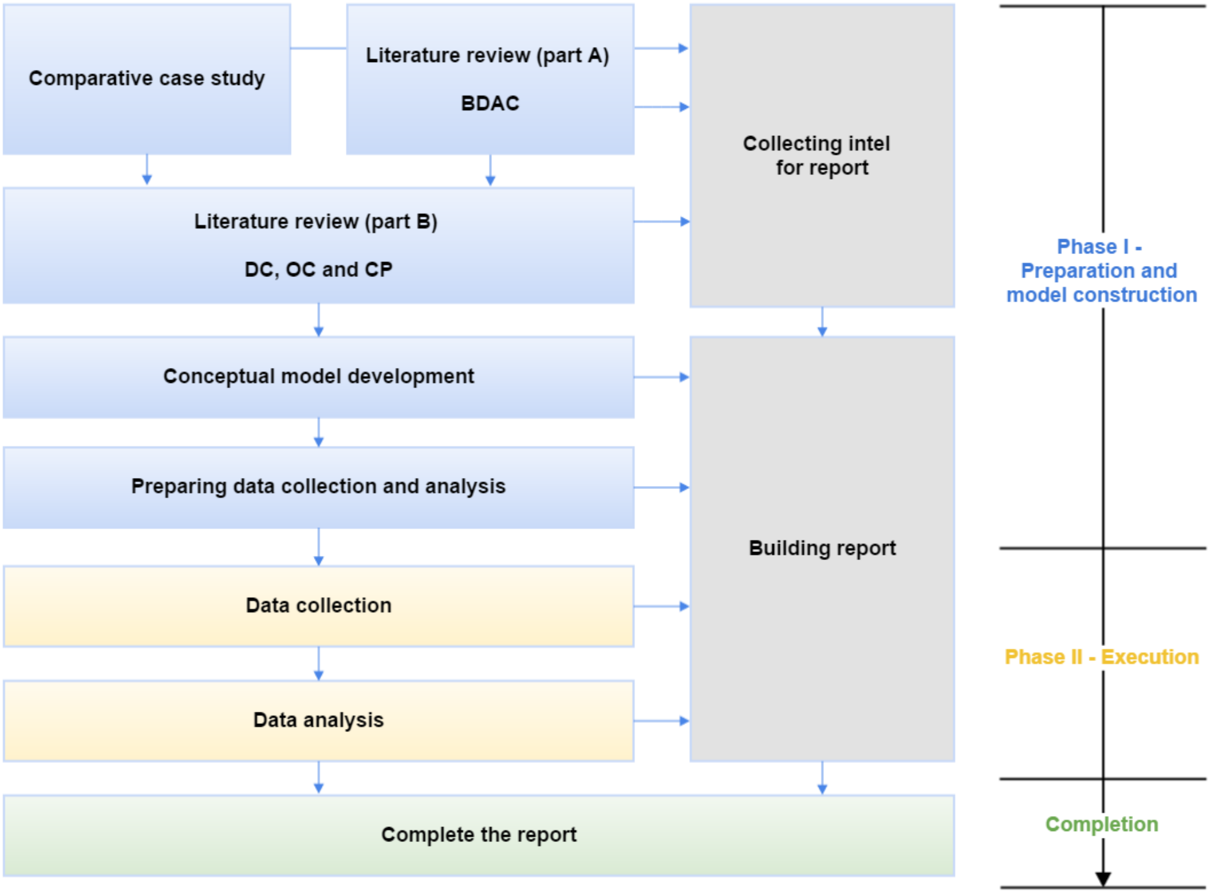


Figure 2: Research design and strategy

Project planning timeline

To successfully conduct a research project of this magnitude, a plan needs to be made with a list of milestones. The plan is an important component to maintain control of all components of the research, from preparations and sub-studies to analysing the results and preparing the discussion goals. This implies constructing and following a timeline with milestones and goals. This also contributed to ensuring steady progress while providing time for reviewing, rewriting and quality assurance. See figure 3 for our plan.

Project planning timeline

Tasks:	40	41	42	43	44	45	46	47	48	49	50	51	52	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Phase I	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X															
Phase II																				X	X	X	X	X	X	X	X	X	X	X	X	X		
Completion																																	X	X
Comparative case study	X	X	X	X	X	X	X	X	X	X	**X**																							
Literature review part A	X	X	X	X	X	X	X	X	X	**X**																								
Literature review part B											**X**	X	X	X	X	**X**																		
Research question(s)														X	X	X	X																	
Conceptual model development															X	X	**X**																	
Designing hypothesis																X	X																	
Improve quantitative research skills																	X	X	X															
Choose survey and analysing tool																		**X**																
Create/alter questions/survey																		X	**X**															
Pilot study																			X	**X**														
Finding respondents																					X	X	X	X	X	X	X	X						
Contact respondents																					X	X	X	X	X	X	X	X						
Send out survey																					**X**	X	X	X	X	X	X	X						
Collect survey																							X	X	**X**	X	X	**X**	X	X	**X**			
Deadline survey 1																									**X**									
Deadline survey 2																												**X**						
Deadline survey FINAL																															**X**			
Analyze survey																												**X**	X	X	X	**X**	x	x
Report progress:																																		
Layout report																X															X	X	X	X
Forword																															X	X	X	X
Abstract																															X	X	X	X
Introduction																			X												X	X	X	X
Key consepts																			X					X	X	X	X				X	X	X	X
Previos research																			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Theory																			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Research approach																								X	X	X	X				X	X	X	X
Results and analytics																												X	X	X	X	X	X	X
Discussion																													X	X	X	X	X	X
Conclusion																															X	X	X	X

BOLD = Milestones

Figure 3: Timeline plan for this research

4.3 Phase I: Preparation and model construction

In this section, we explain the key parts of the process that have shaped our understanding and provided a foundation for the quantitative study. First, we explain the procedure for our case study and the systematic literature review. Next, we explain how the survey questionnaire was designed and how the concepts are defined and operationalized. Lastly, we explain how we secured reliability and validity to the questionnaire.

4.3.1 Exploratory case study process

In parallel to the thorough literature study, an exploratory case study was conducted as a part of this master's thesis. It is sometimes necessary to identify elements that is considered important in the quantitative research through the use of another research method (Oates, 2006, pp. 35, 36). In our case, it was necessary to investigate organizations' use of *big data* and learn how they utilized *big data analytics* and the value added by the practice. By searching for people who worked with *big data* via the business- and employment-oriented social networking service, LinkedIn, we got in touch with several organizations. We selected six interview subjects who were willing to be interviewed. These interviews were conducted face-to-face in five of the six cases. The last one, an organization in Italy, was performed through Skype.

The interviews lasted approximately one hour and were based on a semi-structured interview guide (Oates, 2006, p. 188). This guide was prepared with the aim to provide us insights on the actual use of *big data analytics* in organizations. Specifically, we were interested in exploring: Data (e.g. type of data, durability of data and quality of data), Technology (e.g. analytical tools and advanced data visualization (ADV) tools), Organizations (e.g. involved roles in the *big data* initiative, top management support, data-driven culture, challenges and strategies) and Performance (e.g. potential benefits, different usage and ability adopt to new insight). The semi-structured interview guide developed and matured for each interview as we found interesting results.

The interviews were transcribed and then sent to the interview subjects for validation. This also provided an opportunity to add changes or introduce new responses and comments. When all were approved, we analyzed our transcripts with the aim to trance communalities and differences among interviewees and overall, with the aim to advance our understanding of the domain. This analysis process was based on Creswell (2014, p. 247)'s suggested procedure. Since it was limited to six interviews the coding of the interviews was done manually and not through software tools like, for instance, Nvivo. Then, themes and descriptions were sorted and examined for interesting findings and similarities between the different interviews.

4.3.2 Literature review process

Reviewing prior literature of relevance is a key component in any academic project. By conducting cumulative research, theory building is facilitated. Additionally, research areas that are well-studied are further validated and areas where more research are needed is uncovered (Webster & Watson, 2002). The backbone of this thesis is therefore deeply rooted in prior research. The amount of literature and scientific papers out there is overwhelming. Thus, ensuring sound research at the literature review stage is difficult. By managing the literature in a systematic manner, the review phase is feasible and possesses a high level of quality. Such a systematic approach also helps to avoid so-called “cherry picking” of articles, where the researcher picks literature that speaks in favour of personally established ideas.

There are several papers with pragmatic guidelines to carry out a systematic literature review. Inspired by such papers, a proper systematic approach was chosen and applied to our literature review. The steps of the review method are documented below.

Planning

There are two important factors to consider when searching for articles. Firstly, the sources that are to be used should be determined (e.g. which conferences and journals to include). As the major contributions are likely to be found in the leading journals (Webster & Watson, 2002), the process of collecting articles for both parts of the literature review were aimed towards the “Senior Scholar’ Basket of Journals” also known as “basket of eight”. This is a well-known compilation for the field of information systems, consisting of eight high quality journals (Association for Information Systems, 2011). The electronic databases Google Scholar and Oria were used to trawl the different journals for articles. An overview of the included journal can be seen in table 1.

Table 1: The basket of eight

The basket of eight	Abbreviated title
European Journal of Information Systems	Eur. J. Inf. Syst.
Information Systems Journal	Inf. Syst. J.
Information Systems Research	Inf. Syst. Res.
Journal of AIS	J. Assoc. Inf. Syst.
Journal of Information Technology	J. Inf. Technol.
Journal of MIS	JMIS
Journal of Strategic Information Systems	J. Strateg. Inf. Syst.
MIS Quarterly	Manag. Inf. Syst. Q.

Secondly, one must ensure that the search phrases used yields articles of relevance for the topic of interest. Based on the research question, the phrases were carefully selected by listing out possibilities and performing trial and error searches. To ensure the quality of the search phrases, they were also evaluated in terms of relevance prior to the initiation of the search process. The search strings used for the two parts of the review are shown with explanations in table 2.

Table 2: Literature review search phrases

	Literature review part A (BDAC)		Literature review part B (DC, OC and CP)	
No	Search String	Explanation	Search String	Explanation
1	Big data "analytics capability" OR "analytics capabilities"	We want to look at analytics capabilities.	"Big Data" AND ("Dynamic capability" OR "Dynamic capabilities")	We want to look at dynamic capabilities related 'big data'.
2	Big data capability OR "Big data capabilities"	We want to look for Big Data capabilities.	"Big Data" AND "Competitive performance"	We want to look at competitive performance related 'big data'.
3	Big data strategy OR "Big Data strategies"	There might be good articles that looks at Big Data strategies and therefore also look at the capabilities they have or want to achieve.	"Business Intelligence" AND "Competitive performance"	There might be articles that adresses business intelligence in general and at the same time looking at competitive performance.
4	Big data "business transformation"	This is an interesting combination as we hope to find relevant literature around the adoption phase.	"Business Intelligence" AND ("Dynamic capability" OR "Dynamic capabilities")	There might be articles that adresses business intelligence in general and at the same time looking through the dynamic capabilities view.
5	Big data "business value"	Maybe by looking at articles that talk about the business value of Big Data there can be research into "how" to get these values.	"Business Intelligence" AND "Business value"	A way to understand how 'big data' becomes business value might be through the link between business intelligence and business value.
6	Big data "enabler"	What "enables" a business to adopt or use Big Data?	"Competitive performance" AND ("Dynamic capability" OR "Dynamic capabilities")	How can dynamic capabilities lead to competitive performance?
7	Big data "management"	We search for articles that explains what is needed for Big Data adoption and day to day operations with Big Data.	"Dynamic capability" OR "Dynamic capabilities" (in header)	To understand all aspects of dynamic capabilities this phrase is necessary.
8	Big data "success factor" success factor	What has been done good and right? Is there any articles about this?	"Big data" (in heading)	In general, all articles with Big data in their heading might be important to the research theme.
9	Big data "maturity"	Maybe articles that describes the maturity stages of Big Data might shed light on capabilities.	"evolutionary fitness" (in heading)	The idea that 'big data' can strenghten competitive performance might be described as this phrase.
10	Big data "Performance"	There might be link to "what leads to performance".		

Searching

After having compiled sets of search phrases, they were iteratively used to retrieve published articles from the specified outlets. At this stage, all the search results were included to accumulate a relatively complete census of the relevant literature. This led to a plethora of gathered articles. In figure 4, we provide information on the number of articles identified in the two parts of our literature review and the way we processed them. All the gathered articles were fed into tables with attributes such as name, author, year, journal and which search string that generated the article.

Literature selection

By using this collection method, it is conceivable that some of the search results are not relevant to the research question, even if the search phrases are present in the text of the article. Therefore, the collected articles were manually reviewed to assess the relevance. To facilitate this filtering process, we developed a set of inclusion and exclusion criteria for both parts of the literature review. Articles that met one of the defined inclusion criteria were included as primary studies. On the other hand, the articles that met one of the exclusion criteria were excluded from the study. The inclusion and exclusion criteria are shown below.

In part A, peer-reviewed articles on the following topic were included:

I1: Articles that focus on *big data* and relate to the research question.

Furthermore, articles on the following topics were excluded:

E1: Duplicate articles (multiple entries of the same article).

E2: Articles with a solely technical focus.

E3: Articles that mention the search phrases, but in a context that cannot be related to our research question.

In part B, peer-review articles on the following topics were included:

I1: Articles dealing with a thematic scope that can be linked to our research question.

Articles on the following topics were excluded:

E1: Duplicate articles (multiple entries of the same article).

E2: Articles published before 2002.

E3: Articles that mention the search phrases, but in the context of what cannot be related to our research question.

Unlike part A of the literature review, which exclusively dealt with articles about *big data*, we also considered research that did not reflect on this theme in the second part. By doing so, we chose a more concept centric approach to gather literature. This allowed us to also see how related 'business intelligence' (BI) solutions affected performance which could be useful in shaping hypotheses.

Due to the youth of the field of *big data*, there was no need to incorporate any time constraints in part A of the literature review. The literature on *dynamic capabilities, operational capabilities* and *competitive performance* is on the other hand larger and stems from a more mature field. It was therefore considered necessary to refine the time period in which the articles were published. A timespan of 15 years was selected as appropriate to get a sufficient portion of the literature in the field. The year 2002 thus became a delimiter for exclusion.

The filtering process for both parts (A and B) of the review started out by removing duplicate articles. In part A, this resulted in 67 removed articles and in part B, 45 articles were removed. From there, we scanned and evaluated the relevance of the article title. 50 articles were removed from part A and 49 were removed from part B. Shortly thereafter, we read through the abstract of the articles. At this stage, another 51 articles were peeled off part A and 140 off part B. In the final step of the filtration process, we read through the remaining articles, which resulted in 14 removed articles from part A and 36 removed from part B. Finally, there were eight remaining articles that proved to be relevant in part A of the literature review and 15 remaining in part B. An overview of the filtering process can be seen in figure X.

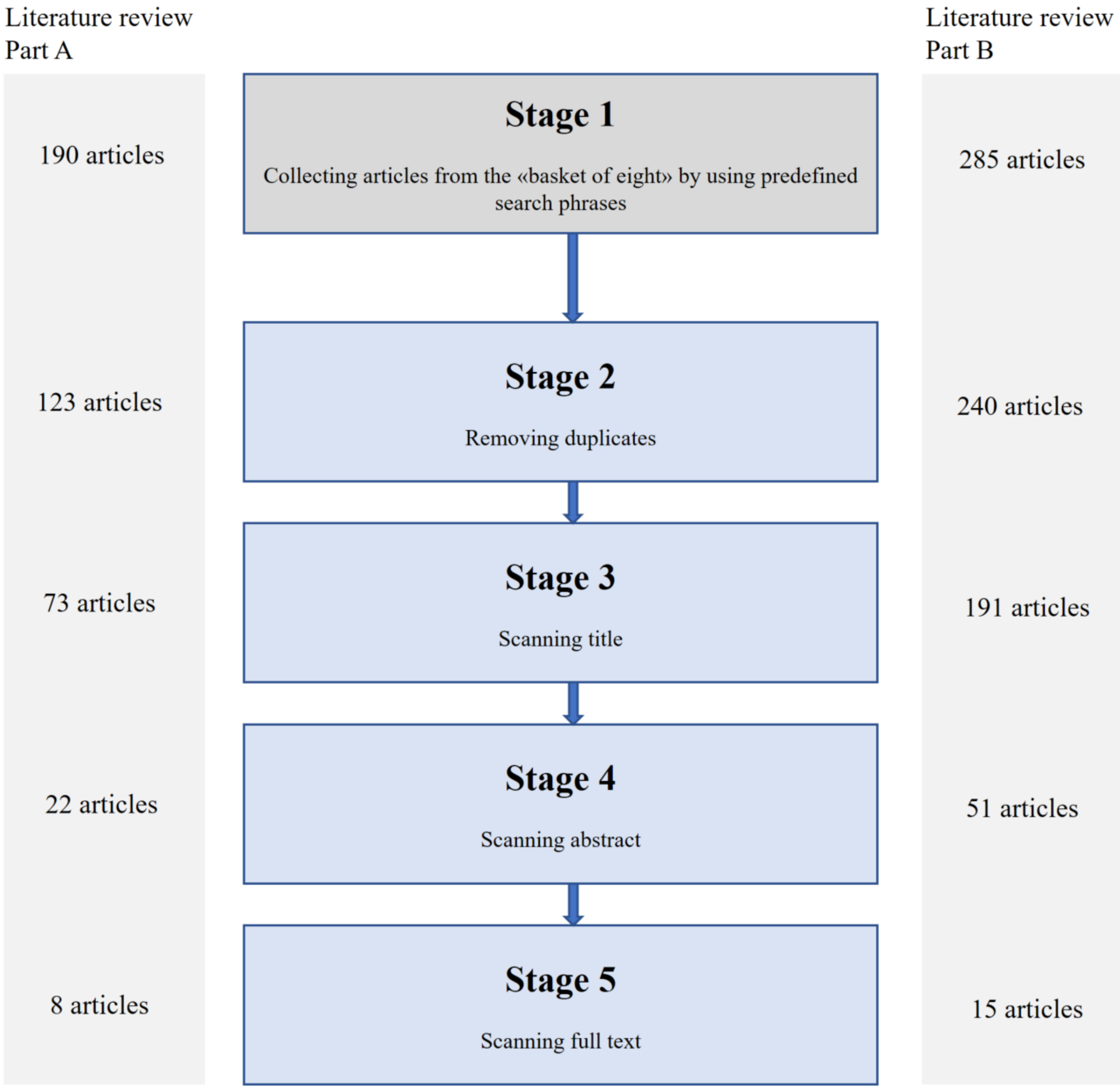


Figure 4: The different stages of our article collection phase

As pointed out by Wamba et al. (2015), the majority of publications in the field of *big data* is, not surprisingly, geared towards the technical aspects as the concept is increasingly viewed as a technology concept. This could be a possible explanation for the low number of relevant articles that was extracted in the filtering process in part A of the literature review. As a consequence of the small number of articles, we saw it as necessary to extend the article pool by doing forward and backward searches, which is a legitimate way to recoup relevant articles (Webster & Watson, 2002). By doing so, the first article pool was expanded with 3 articles.

During the research period, our supervisors, who is as close we get to having an expert panel, supplemented another 26 (nine in part A and 17 in part B) articles which they believed would add value to our study. This is, according to Webster and Watson (2002) also a valid method to retrieve articles.

Collectively, part A of the literature review consisted of 20 peer-reviewed articles and part B consisted of 32 peer-reviewed articles. A final overview of the included articles can be seen in the tables (table 26 and 27) in Appendix 1.

Article distribution by year

The included articles are from the period between 1997 and 2017. Note that the articles supplied by the supervisors are not evaluated according to the exclusion criteria 2 (E2). We have therefore included one article that is older than 15 years due to its central role in the literature stream. As illustrated in figure 5, the articles from part A are mainly during the last four years while the articles from part B are more spread, but with the majority of them published after 2010. Comparing the distributions from both parts of the review shows clear indicators of the freshness of the *big data* phenomenon.

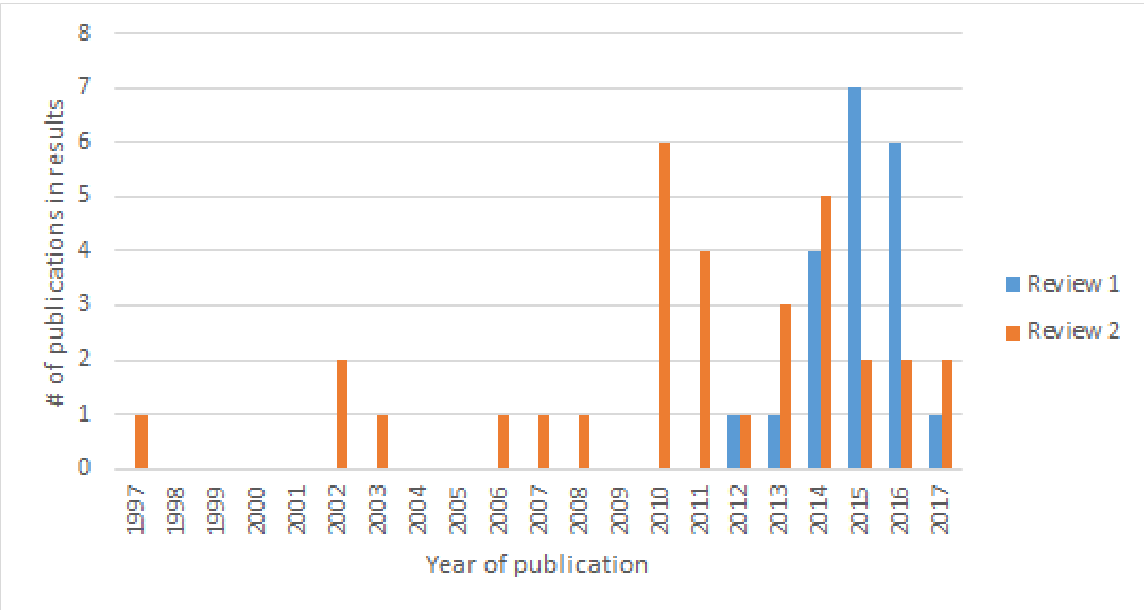


Figure 5: Article distribution by year

Classification

As stressed by Webster and Watson (2002) and echoed by multiple other researchers (Vom Brocke et al., 2009; Wolfswinkel, Furtmueller & Wilderom, 2013), a good literature review should be concept-centric, meaning that the review should be subdivided into concept related units for further analysis. Consequently, we developed a concept matrix with logical partitions to synthesize the studied literature in both parts of the literature review (the matrixes can be seen in table 28, 29 and 30 in Appendix 2). Part A of the literature review was focusing on the concept of *big data analytics capabilities*. To make the matrix, we based our division on Wamba et al. (2017)'s established classification of the concept which is divided into ‘infrastructure flexibility’, ‘management capabilities’ and ‘personnel expertise capability’. When reading through the articles, we became aware of several other capabilities or assets that seemed important to constitute the concept of *big data analytics capabilities*. Constituents like data-driven culture, top management support, data visualization, organizational agility and the organization's analytical capability also seems to affect the firm's ability to use *big data* solutions. In this thesis, we have therefore chosen to use Gupta and George (2016)’s concept classification as it covers these areas and corresponds to our perception of *big data analytics capabilities*. This classification is further divided into ‘tangibles’, ‘human skills’ and ‘intangibles’.

Part B of the review was carried out to cover the concept of *competitive performance* and the paths that connects it to *big data analytics capabilities*. Initially, we looked to the ‘resource-based view of the firm’ (RBV) theory as a theoretical lens to explain this link. But as we came across some of David Teece’s work on the competing concept of *dynamic capabilities*, it became evident to us that this theory was a potentially more suitable lens. Whereas RBV aims to explain “what” resources that increases the company’s *competitive performance*, *dynamic capabilities* aims to explain “how” these resources are developed and integrated (Wade, 2014). The closely related concept of *operational capabilities* was also put under the microscope in part B. This concept was included as a potential link between *big data analytics capabilities* and *competitive performance* because literature from part A of the review pointed out that *big data analytics* bears the potential to streamline work processes by replacing or supporting human decision making with automated algorithms (Wamba et al., 2015). We also found support for this in our exploratory case study. Throughout the review, we also considered enablers and inhibitors such as environmental factors as they appear to be essential in the context of *dynamic capabilities*. In that sense, we chose to adopt the three environmental factors: *environmental hostility*, *environmental dynamism* and *environmental complexity*.

Regarding conceptual breakdowns of the multi-dimensional concepts in part B, there were several different breakdowns that appeared in the literature. In this thesis, we chose to use the breakdowns that best embraced the concept from our understanding and point of view.

From early in the literature stream of *dynamic capabilities,* Teece et al. (1997) breaks the concept into coordination/integration, learning and reconfiguring. This breakdown was later overridden by Teece (2007), who disaggregated the concept into the ability to sense opportunities, to seize opportunities and to reconfigure the firm’s asset orchestration. This classification is also rendered by Roberts et al. (2016) and Wilden et al. (2013). Mikalef and Pateli (2017) breaks the concept into sensing, coordinating, learning, integrating and reconfiguring. We chose to adopt the breakdown of Teece (2007) as we saw it to be comprehensive.

As of the concept of *operational capabilities*, Wu et al. (2010) divides the concept in to the six following capabilities: operational improvement, operational innovation, operational customization, operational cooperation, operational responsiveness and operational reconfiguration. Another division is done by Wilden and Gudergan (2015), which divides it into marketing and technological capabilities. As the first breakdown had some overlap with the concept of *dynamic capabilities* and as we saw the second one as appropriate, we chose to adopt the multi-ordered construct of Wilden and Gudergan (2015).

4.3.3 Online Survey Design

As we only had one opportunity with our respondents, a lot of effort was put in preparation for the survey. Our supervisors had a crucial role in ensuring that our constructs were properly measured and that the questions were formulated in an easy-to-understand manner so we could generate the data needed to test our hypotheses and answer our research question.

To make the survey, the survey software tool SurveyGizmo (SurveyGizmo, 2017) was used. The survey was designed so that respondents were first met with an introductory text about the context of the study, who supervised the study and an assurance that the data would be treated anonymously. To add validity to the survey, the logos from both UiA and NTNU were used at the front page (see figure 12 in Appendix 3 for image excerpt of the survey).

Although SurveyGizmo was equipped with an email campaign function that allowed bulk mailings, we chose to send out the surveys one by one from our private university email addresses. This was done as we are aware that most email spam filters prevent bulk emails to be delivered in the inbox (Oates, 2006, p. 102) and tests performed through SurveyGizmo confirmed this.

4.3.4 Construct Definition and Measures

To provide answers to our research question, hypotheses and research model, we had to put down extensive effort in finding the right questions (indicators) to measure our model's variables. The literature review contained many examples of earlier, well-established operationalization's of the different variables. Most of the questions we ended up using were from these previously used surveys even if they had different agendas. The way we chose what questions to include was to list all found questions for each variable, including our own, and provide a relevance score to each one. Those with the highest score were chosen and a verification was performed by our supervisors. Chen et al. (2014) also prepared their survey the same way, by using previously validated indicators and if necessary modifying them slightly to fit the new context.

The next sections will explain the operationalization's of our model and the different variables.

Operationalization of control questions

In addition to the questions related to the variables, we provided a few introductory questions. These were included to collect demographic information and support the analysis. We asked if the respondents organization uses *big data analytics*. It was important to secure that the organizations that participated in the survey did use *big data* solutions. In addition to information in the inviting emails we added a yes/no question and a description of how we define *big data*. Also, we wanted to identify how long each participant had used *big data* solutions. Investments made by organizations might not provide value at once. Schryen (2013) suggests it may take years. In our case, where we focus on *big data analytics* investments, it is hard to say when they will see results and since *big data* is a relative fresh phenomenon we chose to measure duration with alternatives in the range of "less than one" and "more than four". Furthermore, age and size should provide good background information on the organizations and perhaps extend our findings. We measured size in accordance with the recommendations enacted by the European Commission (European Commision, 2012), with the following values: micro (0-9 employees), small (10–49 employees), medium (50–249 employees) and large (250+ employees). Lastly, we needed to ask all respondents which industry the organization operate in. If possible, we could identify findings based on different industry sectors. We used the industry-list from Kapital 500 (2016) and added a free-text option in order for respondents to add their own industry if needed. The questions are shown in table 3.

Table 3: Operationalizing demographic questions

Name	**Questions**
BG0	Is your organization using 'big data analytics'?
BG1	When did your organization start using 'big data analytics' solutions?
BG2	How old is your organization? (measured in years)
BG3	Please indicate the size-class of your organization. (Number of employees)
BG4	In which industry does your organization operate? (multiple choice + free-text)

Operationalization of big data analytics capabilities

As stated in chapter 2, the construct presented by Gupta and George (2016) provides a good basis for including assets and capabilities belonging to *big data analytics*. They define *big data analytics capabilities* as a third-order construct and divided the concept *big data analytics capabilities* into tangibles, intangibles and human skills. We adopted this construct and we show it in figure 6.

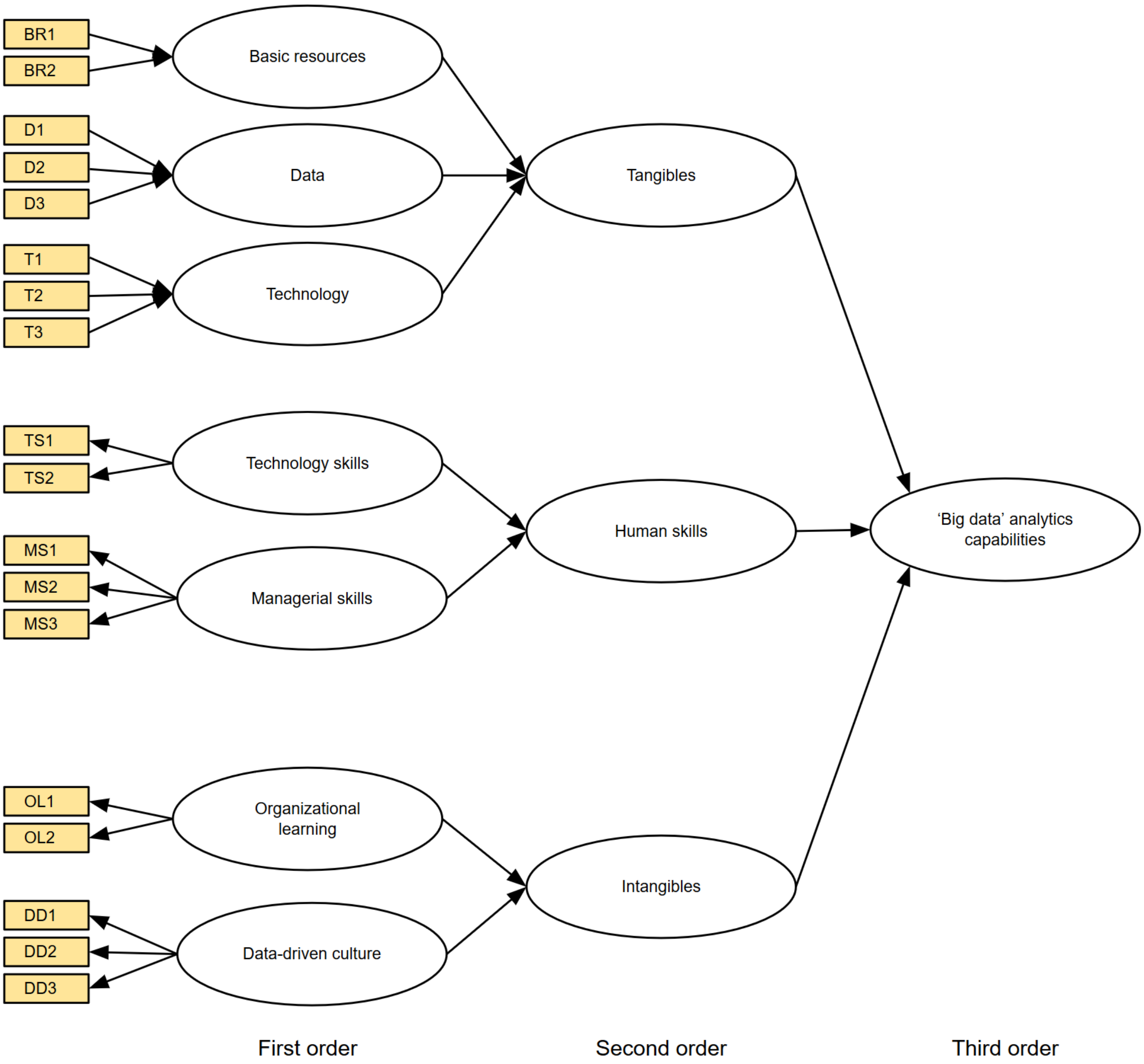


Figure 6: The third-order construct of big data analytics capabilities

Tangibles

Tangibles are divided into basic resources, data and technology. These assets can be sold or bought in a market (Gupta & George, 2016). For instance, they could be financial resources or physical resources. The questions can be viewed in table 4.

Basic resources contain both time and investments (Gupta & George, 2016). That way, organizations can be measured for the strength of their concepts and basic resources when it comes to investing in *big data* initiatives and letting them have enough time to bear fruit.

Data resources are obviously important. Organizations collect data at an increasing speed (Gupta & George, 2016). Connectivity and access to these data from the different business functions strengthen this first-order term (Akter et al., 2016) while strong departmental data strongholds are caused by resistance against sharing and merging data across the organization. Another side of the data resource is compatibility. This refers to the ability to support constant flows of information to enable decisions to be taken at near real-time (Akter et al., 2016). Do organizations have access to *big data*? Do they have traditional or untraditional systems to store those data and do they manage to integrate internal data with external data? These are questions that can measure this first-order term (Gupta & George, 2016).

Technology is also an important part of *big data analytics capabilities*' tangible resources. Organizations that want to use *big data analytics* need to have some type of database management systems. This can be relational (RDBMS) but often it requires untraditional systems since *big data* offer new and difficult challenges. The need for systems (e.g., Hadoop) to process *big data*, new and adjusted visualization tools and new database technologies for data storage is a good way to measure the technology resource, in relation to *big data* and *big data analytics capabilities* (George et al., 2014; Wamba et al., 2017).

Table 4: Operationalizing basic resources, data and technology

Name	Question	Sources
Basic resources	Our 'big data analytics' projects are ________________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
BR1	adequately funded	(George et al., 2014)
BR2	given enough time to achieve their objectives	(George et al., 2014)
Data	Our organization's **data capabilities** can be described by the following statements: *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
D1	We have access to very large, unstructured, or fast-moving data for analysis	(George et al., 2014)
D2	We integrate data from multiple internal sources into a data warehouse or mart for easy access	(George et al., 2014)
D3	We integrate external data with internal to facilitate high-value analysis of our business environment	(George et al., 2014)
Technology	We have explored or adopted ________________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
T1	parallel computing approaches (e.g., Hadoop) to big data processing	(George et al., 2014)
T2	different data visualization tools	(George et al., 2014)
T3	new forms of databases such as Not Only SQL(NoSQL) for storing data	(George et al., 2014)

Human skills

An organizations human skills consist of its employees' experience, knowledge, business acumen, problem-solving abilities, leadership qualities, and relationships with others (Gupta & George, 2016). Earlier research suggests that technical, business and relational knowledge is included in this second-order term in addition to technical management (Garmaki et al., 2016; Wamba et al., 2017). We choose to focus on technical skills and managerial skills as these are important aspects of an organization's *big data* resources (Gupta & George, 2016). The questions can be viewed in table 5.

Technology skills refer to the know-how required to use new forms of technology to extract intelligence from *big data* (Gupta & George, 2016). This could consist of, for instance, knowledge regarding technical elements like operating systems, database management systems (DBMS), programming languages, and statistical analysis (Akter et al., 2016). More specifically, the skills related to machine learning, data extraction, data cleaning, statistical analysis, and understanding of programming paradigms such as MapReduce (Gupta & George, 2016). It can be measured in how well organizations rate when it comes to owning the skills to perform *big data analytics* (with success).

Managerial skills are skills specific to each organization and are developed over time (Mata et al., 1995). Collaboration between the different departments and good working relationships strengthen this set of skills. To measure managerial skills, we asked reflective questions like "do *big data* managers understand an organization's business needs?", "how do *big data* managers collaborate with other managers, customers and suppliers?" and "do *big data* management understand the results *big data* provide?"

Table 5: Operationalizing technology skills and managerial skills

Name	Question	Sources
Technology skills	Our 'big data analytics' staff ________________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
TS1	has the right skills to accomplish their jobs successfully	(George et al., 2014)
TS2	is well trained	(George et al., 2014)
Managerial skills	Our 'big data analytics' managers are able to ________________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
MS1	understand the business need of (and collaborate with) other functional managers, suppliers, and customers to determine opportunities that big data might bring to our business.	(George et al., 2014)
MS2	coordinate big data-related activities in ways that support other functional managers, suppliers, and customers	(George et al., 2014)

MS3	understand and evaluate the output extracted from big data	(George et al., 2014)

Intangibles

Intangible resources are, unlike tangible resources, not documented on an organization's financial statements (Grant, 2016, p. 128). Intangible resources might not be tradeable in a market since they are highly context-dependent. There are exceptions like, for instance, trademarks, copyrights, patents or franchises, which could be tradeable. Gupta and George (2016) divided this into two assets in relation to *big data analytics*; organizational learning and data-driven culture. The questions can be viewed in table 6.

Organizational learning is the ability of an organization to exploit existing knowledge and explore new knowledge (Gupta & George, 2016). Related to the intelligence coming from *big data analytics* and what that might provide of strategic changes for an organization, it is important to include how well an organization can adapt to new configurations. An organization with high intensity of organizational learning will probably get an advantage in extracting value from *big data analytics* results (Gupta & George, 2016) To measure organizational learning, we used reflective questions that focused on how well organizations acquire new and relevant knowledge and how much they exploit existing competencies and explore for new knowledge.

Data-driven culture is the ability to use data in the decision processes. Educating key personnel (or users of *big data analytics* results) to appropriately interpret the results will build up the competency of data-driven decisions (Wang, Kung, Ting & Byrd, 2015). While *big data analytics* is an enabler for improved decision making (Wamba et al., 2015), it will redefine how decision making is done and how it affects authority, influence and organizational power (Bhimani, 2015). Therefore, with the above mentioned in mind, we chose reflective questions that measured at what extent organizations base their decisions on data versus instincts (or intuitions) and how much they educate employees to make decisions based on data.

Table 6: Operationalizing organizational learning and data-driven culture

Name	Question	Sources
Organizational learning	Our organizational **learning capabilities** can be described by the following statements: *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
OL1	We are able to acquire new and relevant knowledge	(George et al., 2014)
OL2	We have made concerted efforts for the exploitation of existing competencies and exploration of new knowledge	(George et al., 2014)

Data-driven culture	Our organization's **data-driven culture** can be described by the following statements: *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
DD1	We base our decisions on data rather than on instinct	(George et al., 2014)
DD2	We are willing to override our own intuition when data contradict our viewpoints	(George et al., 2014)
DD3	We continuously coach our employees to make decisions based on data	(George et al., 2014)

Operationalization of dynamic capabilities

According to (Wilden & Gudergan, 2015), most of the conducted research on *dynamic capabilities* has been theoretical due to the difficulty of measuring its effects. In our literature review, however, most studies on *dynamic capabilities* are conducted on an empirical basis. These studies have different ways to operationalize the concept.

In this study, *dynamic capabilities* was measured as a second-order construct (reflective first-order and formative second-order) consisting of three first-order constructs. The first-order constructs were based on (Teece, 2007)'s conceptualization that is divided into the ability to sense opportunities, to seize opportunities and to transform the organizations resource base accordingly, all of which consisted of three items as illustrated in figure 7. The questions can be viewed in table 7.

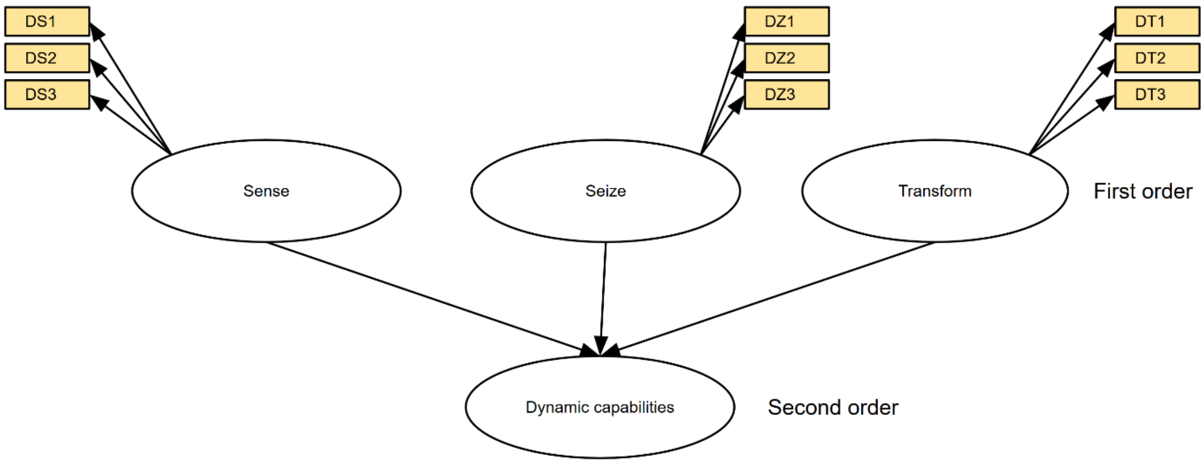


Figure 7: The second-order construct of dynamic capabilities

Sense involves searching for new business opportunities and learning about customers, competitors, and the broader market environment (Teece, 2007; Wilden & Gudergan, 2015). The questions were constructed to measure these areas of investigation.

Seize describes an organizations ability to seize opportunities (Teece, 2007). There were no articles with previous questions on this type of defining *dynamic capabilities*; therefore, questions had to be constructed based on theory. We developed questions on how organizations draft, evaluate or carry out potential solutions when threats or opportunities occurs.

Transform is the process of reconfiguring the organizations resource base (Teece, 2007). According to this definition, we chose to use questions focusing on how organizations have changed to achieve new objectives, adjustments following changed business priorities and changing business processes.

Table 7: Operationalizing sense, seize and transform

Name	Question	Sources
Sensing	Our organization's sensing capabilities can be described by the following statements: *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
DS1	We frequently scan the environment to identify new business opportunities	(Pavlou & El Sawy, 2011)
DS2	We often review our product development efforts to ensure they are in line with what the customers want	(Pavlou & El Sawy, 2011)
DS3	We use established processes to identify target market segments, changing customer needs and customer innovation	(Wilden et al., 2013)
Seizing	When opportunities or threats are sensed, our organization has effective routines for _______________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
DZ1	drafting various potential solutions	(Ortbach, Plattfaut, Poppelbuß & Niehaves, 2012; Teece, 2007)
DZ2	evaluating and selecting potential solutions	(Ortbach et al., 2012; Teece, 2007)
DZ3	starting on a detailed plan to carry out a potential solution	(Ortbach et al., 2012; Teece, 2007)
Transforming	Our organization can successfully carry out the following activities: *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	

DT1	Create new or substantially changed ways of achieving our targets and objectives	(Wilden et al., 2013)
DT2	Adjusting our business processes in response to shifts in our business priorities	(Wilden et al., 2013)
DT3	Reconfiguring our business processes to come up with new productive assets	(Wilden et al., 2013)

Operationalization of operational capabilities

Drnevich and Kriauciunas (2011) describes *operational capabilities* as an organization's ability to "make a living". *operational capabilities* are connected to technical fitness (Li et al., 2010) and marketing capabilities (Wilden & Gudergan, 2015). We chose to adapt the work of Wilden and Gudergan (2015) as a way of defining this capability, which is to divide *operational capabilities* into two subgroups, marketing and technological capabilities. See figure8 for an overview over the construct of *operational capabilities* and its related items (questions). The questions can be viewed in table 8.

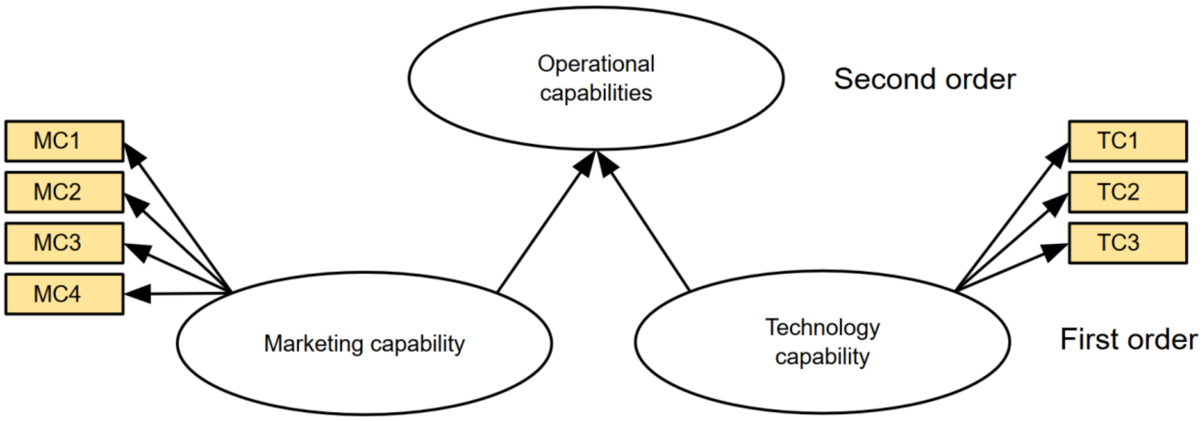


Figure 8: The second-order construct of operational capabilities

Marketing capabilities is the ability of an organization to control and extract value from the market in which operates (Wilden & Gudergan, 2015). To measure these capabilities, we chose reflective questions that asked at what level about the organization controlled market knowledge, customer relationships and distribution channels.

Technological capabilities describe an organization's ability to perform its day to day processes and can be measured in how efficient or effective they are at converting inputs to outputs (Wilden & Gudergan, 2015). The questions we selected, which are reflective, was constructed around this efficiency, handling of day to day operations, technical expertise and equipment.

Table 8: Operationalizing marketing capabilities and technological capabilities

Name	Question	Sources
Marketing capabilities and Technological capabilities	Our organization has excellent capabilities when it comes to ________________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
MC1	Market knowledge	(Wilden & Gudergan, 2015)
MC2	Control and access to distribution channels	(Wilden & Gudergan, 2015)
MC3	Advantageous relationships with customers	(Wilden & Gudergan, 2015)
MC4	Established customer base	(Wilden & Gudergan, 2015)
TC1	Efficient and effective production/services	(Wilden & Gudergan, 2015)
TC2	Economies of scale and technical expertise	(Wilden & Gudergan, 2015)
TC3	Technological capabilities and equipment	(Wilden & Gudergan, 2015)

Operationalization of competitive performance

Competitive performance refers to the degree a firm performs better than its key competitors (Mikalef & Pateli, 2017; Rai & Tang, 2010). To measure the *competitive performance*, we chose reflective questions that focused on profitability, return on investment, growth in market share and sales, reduction of operating costs, increasing customer satisfaction and provisioned rapid response to market demand. See figure 9 and table 9 for an overview of the items (questions) belonging to *competitive performance*.

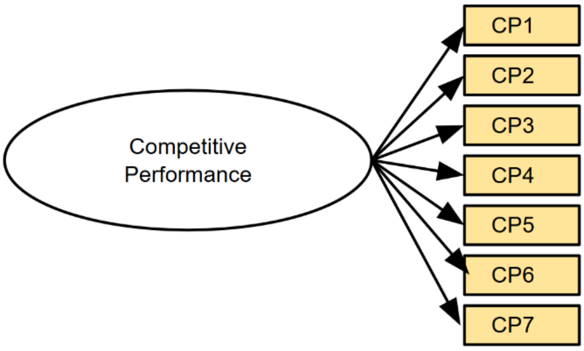


Figure 9: Competitive performance

Table 9: Operationalizing competitive performance

Name	Question	Sources
Competitive performance	We perform much better than our main competitors in _______________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
CP1	profitability	(Chen et al., 2014)
CP2	return on investment (ROI)	(Chen et al., 2014)
CP3	growth in market share	(Chen et al., 2014)
CP4	sales growth	(Chen et al., 2014)
CP5	rapid response to market demand	(Mikalef & Pateli, 2017)
CP6	in reducing operating costs	(Mikalef & Pateli, 2017)
CP7	increasing customer satisfaction	(Mikalef & Pateli, 2017)

Operationalization of environmental factors

The moderating effect of environmental factors on *competitive performance* and organizational capabilities is well supported in the literature (Chen et al., 2015; Chen et al., 2014; Rai & Tang, 2010; Wilden et al., 2013). By looking at the literature and adapting the work by Chen et al. (2014) their division of factors, we chose to measure *environmental hostility*, *environmental dynamism* and *environmental complexity*. These questions were used in earlier surveys and Chen et al. (2014) did extensive work in evaluating these questions before conducting the survey. See figure 10 for an overview of the construct of environmental factors, where they appear as moderators (they moderate on H2 and H4) and they are linked to their related items (questions). The questions can be viewed in table 10.

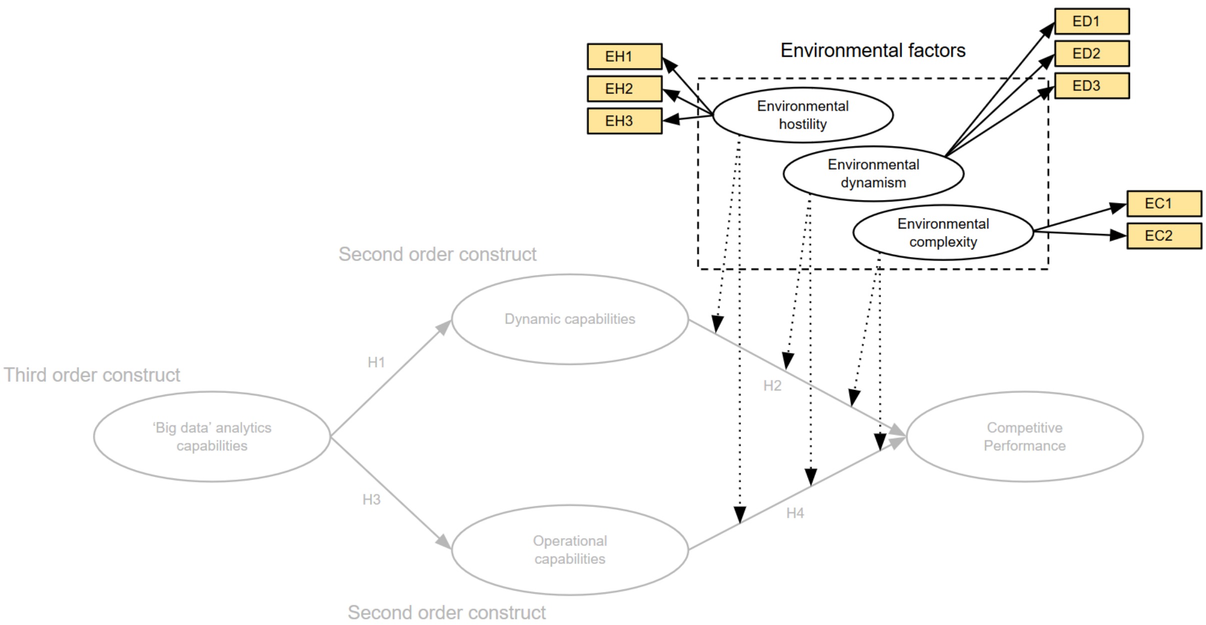


Figure 10: The environmental factors

Environmental hostility describes if an (Chen et al., 2014). To measure this environmental factor, we asked about competition when it comes to pricing, quality and the differentiation of product or services.

Environmental dynamism describes if an environment is unpredictable. Earlier literature measures this by asking reflective questions concerning the rate at which products and services become obsolete. We also asked, at what rate the technologies were associated with products and service changes. We also asked about the rate the competitors change their behaviours (Chakravarty, Grewal & Sambamurthy, 2013; Chen et al., 2014).

Environmental complexity describes if an environment consists of complex external factors. Finite resources and competition might strengthen (making it more complex) this type of environment. With support from earlier research, we chose to measure *environmental complexity* by asking reflective questions about customer buying habits and the nature of competition that exists in the organization's environment (Chen et al., 2014).

Table 10: Operationalizing environmental factors

Name	Question	Sources
Environmental complexity	In our industry, there is considerable diversity in ________________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
EC1	customer buying habits	(Chen et al., 2014)
EC2	nature of competition	(Chen et al., 2014)
Environmental hostility	The survival of our organization is currently threatened by tough ______________ *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
EH1	price competition	(Chen et al., 2014)
EH2	competition in product/service quality	(Chen et al., 2014)
EH3	competition in product/service differentiation	(Chen et al., 2014)
Environmental dynamism	The environmental dynamism our organization operates in can be described by: *Please rate each statement based on the level you agree or disagree* ***(1 - totally disagree, 7 - totally agree)***	
ED1	In our industry, products and services become obsolete quickly	(Chen et al., 2014)
ED2	The product/services technologies in our industry change quickly	(Chen et al., 2014)
ED3	Our competitors' behaviors exhibit a lot of variability	(Chakravarty et al., 2013)

Validity of the questionnaire

We followed Oates (2006, p. 227) recommendations for securing the reliability and validity of our questionnaire. Our strategy was that investigation into what other peer-reviewed articles did to ensure this lead to well established methods we could imitate. Also, by choosing earlier tested measurements from these articles we could include their history of refinement and therefore secure content validity. In addition, the usage of our supervisors' expertise as another quality insurance further refined the questionnaire.

Content validity

To secure content validity we had to make sure the questions were well balanced in the domain to be covered (Oates, 2006, p. 227). The way we accomplished this was to use prior literature and questions from questionnaires used in other surveys, all from our literature review. All questions from research surveys that measured our different latent variables were collected and thoroughly evaluated before we selected the best ones. This was done by rating the questions on a 1-10 scale and by discussing each question. We then determined which one best describes the variable. We had a pool of 278 questions which was downsized to 67. This was followed by a process where our supervisors evaluated both our finished survey and our rating on the questions and reasoning for choosing as we did.

Construct validity

Oates (2006, p. 227) says that construct validity is concerned with whether we are measuring what we think we are measuring. Some questions might measure other sides of our research model than what is intended. Due to the limited number of potential respondents to our survey, we decided not to pilot test our constructs (to avoid further reducing our pool of respondents to the main survey). Hence, we were not able to check correlations before sending out the survey. Nevertheless, we tried to check this as much as we could by examining other surveys from the literature and learn what they struggled with and carefully choose the questions.

4.4 Phase II: Execution

In this section, we explain the methods used in the second phase of the study. This phase entails the method used for the survey data collection and analysis. Lastly, we will go through how we secured reliability and validity for the data analysis.

4.4.1 Method for collecting data

In this part, we explain the procedure we used to define our target population, the sampling techniques we used, the construct operationalisations and the survey design.

Population and sampling frame

As implied by the research question, we were interested in organizations that had adopted analytical tools to handle *big data*. We were aware that the concept was fairly new in the business world and that only a very few organizations had yet adopted *big data analytics* solutions (Kwon et al., 2014). We were therefore also aware that to locate these organizations would be a major challenge.

For us to be able to perform reliable statistical analysis of collected data, we were dependent on getting enough respondents. The required sample size for reliable statistical analysis depends on the type of methods to be employed. For our research, we had decided to use structural equation modelling (SEM). Specifically, we used PLS-SEM as explained in section 4.4.3. Therefore, we needed a minimum sample size equal to or bigger than the largest of the following two requirements:

"10 times the largest number of formative indicators used to measure a single construct, or

10 times the largest number of structural paths directed at a particular construct in the structural model." (Hair Jr, Hult, Ringle & Sarstedt, 2013, p. 20)

In our conceptual model, there are three constructs with formative indicators. Those are Basic resources (two indicators), Data (three indicators) and Technology (three indicators). These are part of a multi-ordered structure and therefore Tangibles have 8 formative indicators inherited from these first-order constructs. The largest number of structural paths directed at a construct is 3, which points to *big data analytics capabilities*. This meant that we had to have a sample size of at least 80 (10*8=80). As pointed out by (Wong, 2013), merely fulfilling the minimum sample size requirements should not be the goal. A sample size of 100 to 200 is suggested as a good starting point. In consultation with our supervisors, we agreed to aim for a sample size of 100+ to do sound research.

Based on our conception, and after consulting with our supervisors, we agreed to aim the study towards companies in the Nordic region (Norway, Sweden, Denmark, Finland and Iceland) as this would raise our chances of getting enough respondents. Although contextual factors at the country level may impact the link between IS investments and the outcome of firm performance (Schryen, 2013), we believe that the differences between the countries are minor as all companies operate within the Nordics. According to World Economic Forum's Global Information Technology Report (GITR) from 2016, all the Nordic countries rank high on the list (Baller, Dutta & Lanvin, 2016).

As our topic of interest covers both technical and business aspects, we wanted to retrieve data from representatives that could have insight into both aspects. The survey was therefore primarily directed to executives with roles such as Chief Information Officers (CIO), Chief Technology Officers (CTO), Head of *big data* -, analytics- or business intelligence department.

Sampling technique and contacting respondents

Due to uncertainty, related to the size of our population and limitations related to research resources, we chose to use several non-probability sampling techniques (Oates, 2006, p. 97). Because many of the pioneers of *big data analytics* were large and recognized internet companies (Chen et al., 2012), we assumed that turning to large companies would yield a higher response rate. We thus chose to use a purposive sampling technique where we at first focused on companies on the Norwegian Kapital 500 list and the Nordic companies appearing on the Forbes Global 2000 list, which both are lists of large companies that might be part of our population (Forbes, 2017; Kapital 500, 2016). We also did more targeted searches by looking through LinkedIn, *big data* "meetups" and job advertisements seeking *big data* expertise

(Finn.no, 2017; Finnson, 2017; Indeed, 2017; Jobbsafari, 2017; Meetup, 2017). As we came across companies that potentially matched our population, we added them to an internal collecting document that gave us an overview of companies to be contacted.

The contacting activity went on for about eleven weeks on a daily basis (from 20.02.2017 to 07.05.2017). Initially, we contacted companies and respondents only by phone. This proved to be far more resource-intensive than what we first thought, especially in large companies where IT management was decentralized or outsourced. We therefore changed our contacting strategy to find representative respondents from the population either via LinkedIn or browsing the web and then send them mail directly. This enabled us to contact more respondents per day. Finding the right mail addresses was done in various ways, such as searching the web, using the e-mail lookup software RocketReach (RocketReach LLC, 2017) or simply guessing the address.

We also used two types of snowball sampling techniques (Oates, 2006, p. 98) where (1) LinkedIn suggested similar companies and several of (2) contacted respondents referred to other potential respondents that could be of interest. This led to a further growth of the sample, which eventually ended up being 557 companies. We contacted all those companies, requesting them for participation in the survey.

As an incentive, we offered all participant that completed the survey a personal benchmark that showed their responses on the survey, compared to the average answers. To further sweeten the deal, we offered a copy of our final master thesis report. The plan was to deliver these reports to the respondents in June 2017.

One respondents per organization was invited, in some cases where we did not get any answers from the invited manager, we invited another manager if this was a possibility. Many of the biggest organizations we contacted had several IT managers, business intelligence managers et cetera.

To increase the response rate, two rounds of reminders were sent to all respondents who had not completed the survey.

4.4.3 Method for analysing data

Univariate and bivariate analysis has for long been the statistical method of choice to study data and its correlations. The rapid technological development has given rise to far more sophisticated analysis methods and tools that have made it possible for researchers to study far more complex relationships (Hair Jr et al., 2013, p. 2). In this thesis, we applied such a sophisticated method called Partial Least Squares Path Modelling (PLS-SEM), which is a type of Structural Equation Modelling (SEM). PLS-SEM constitutes one of the two types of SEM. The other type is called Covariance-based SEM (CB-SEM) and is mainly used for testing established theories. PLS-SEM on the other hand is recommended for performing tests in circumstances where theory is less developed and when the structural model is complex (Hair Jr et al., 2013, pp. 14,19).

In an effort to raise our expertise in quantitative, PLS-SEM methodology, we travelled to NTNU in Trondheim for a two-day workshop (from April 27 to April 28th). This was held by our supervisor Patrick Mikalef who has applied this method in previous research.

To perform the data analysis, we used the analysis software called SmartPLS (Ringle, Wende & Becker, 2015), which allowed us to visually create a path model that matched our conceptual model with our variables and hypotheses.

To ensure that the PLS-SEM analysis was performed on the correct basis, the collected data underwent a thorough cleansing. First, a total of 135 responses were extracted from SurveyGizmo and manually encoded to a numeric format. Thereafter, 28 responses were filtered out as they stated that they did not use *big data analytics* according to the provided definition. This left us with 107 usable answers that were fed into SmartPLS.

Reliability and validity: Analysing

By performing PLS calculations, we performed an evaluation of the conceptual model. Based on the sample data and the structural model we determined how well the conceptualization fits with reality. There are many potential sources of measurement errors when conducting surveys on people (such as poorly formulated questions, misinterpretation of Likert scales or incorrect use of statistical method). Consequently, great effort were made to remove as many measurement errors as possible (Hair Jr et al., 2013, pp. 96, 97).

To ensure that we performed a proper evaluation of the model and to remove as many error sources as possible, we made an action plan (table 12 and 13). The action plan was based on two separate evaluations. First, an evaluation of the outer model, meaning the indicators and the relationships that connects them to their intended measured factors (Garson, 2016, p. 60). Second, an evaluation of the inner model, meaning the paths between the latent variables (Hair Jr et al., 2013, p. 116). Our conceptual model has several higher order constructs which are of the type *reflective-formative,* meaning that the first-order constructs are outlined as reflective (arrows point from the first-order construct to the indicators) and the second-order constructs are outlined as formative (arrows point from first-order to second-order construct) (Garson, 2016, p. 236) (see figure 14 in Appendix 4 for an extensive overview of the measurement model).

To ensure that we check reliability and validity in the best possible way in terms of our analytical results, we have used a variety of sources. Textbooks like Oates (2006) and Hellevik (2011) introduced us to these processes. Reading master's theses that were related to quantitative methods are also a good source of inspiration. The literature review provided us with in-depth information about similar methods of research and therefore these was of great help and inspirations when planning how to secure good and valid results which, not only established procedures but also new solutions that are not mentioned in the textbooks. Lastly, the book written by Hair Jr et al. (2013) has been diligently followed. This is a book that focuses on PLS-SEM and SmartPLS and therefore a good source. We noticed that several articles refer to this book in their arguments (Gupta & George, 2016; Wamba et al., 2017). The rules of thumb presented in this book that were followed is presented in table 11.

Table 11: Reliability and validity, rules of thumb

Analysing the measurement models	
Formative measurements	Reflective measurements
Convergent validity	Indicator reliability
Collinearity among indicators	Composite reliability
Significance and relevance of outher weights	Convergent validity (AVE)
	Discriminant validity
Analysing the structural model	
Collinearity Assessment	
Structural Model Path Coefficients	
Coefficient of Determination (R2)	
Effect Size (f2)	
Predictive Relevance Q2 (and q2)	

Analysing the measurement models (Outer model)

The outer model consists of all measurements of the latent variables. In our case, we used multi-ordered constructs and therefore we include the measurements of the second and third order constructs in this section. We followed the premade action plan for outer model analysing, which can be viewed in table 12.

Table 12: Action plan, outer model

Outer model	Formative measures	Significance and relevance of outer weights	T-values
			P-values
		Convergent validity	Erstattes av Adequacy coefficient
		Collinearity among indicators	VIF (Variance Inflation Factor)
		Adequacy coefficient	R2a
	Reflective measures	Reliability	Composite reliability
			Cronbach's alpha
			Indicator reliability
		Convergent validity	Average variance extracted (AVE)
		Discriminant validity	Cross loadings
			Fornell-Larcker criterion
			Heterotrait-monotrait ratio (HTMT)

Formative measurements

To evaluate the reliability and validity of the outer model we first evaluated the formative measurements by looking at the significance and relevance of outer weights. This was done by looking at the calculated t-values and transfer them to p-values. To do this we used bootstrapping in SmartPLS and predefined table (Hellevik, 2011; Lind, Marchal & Wathen, 2007). Since there are elements in our model that could show negative values, which also is interesting in itself, this t-value calculation were run as a two-tailed (two sided) test (Hair Jr et al., 2013, p. 172; UCLA, 2017). Variance Inflation Factor is measured to check if the formative constructs represent multicollinearity. For formative measures (Petter, Straub & Rai, 2007) suggest values below 3.3 are low multicollinearity. Further, to evaluate the validity of the formative measurements on the constructs, adequacy coefficient (R^2_a) were used (Edwards, 2001). This is not calculated in SmartPLS and the way to do this is to sum the squared correlations between the formative measurements and the constructs involved and then divide the sum by the number of measurements. These values should be above 0.50.

Reflective measurements

Evaluating the reflective measures is done by techniques other than those used for formative measures. With the help of SmartPLS and our own calculations we followed the rule of thumb presented by (Hair Jr et al., 2013, p. 107) and checked reliability, convergent validity, and discriminant validity. First, we measured for reliability by examining composite reliability, Cronbach's alpha and indicator reliability. Composite reliability and Cronbach's alpha should have values above 0.708 (Hair Jr et al., 2013, p. 115; Nunally & Bernstein, 1978) and each reflected indicator should have a loading above 0.708 (Hair Jr et al., 2013, p. 109). Then we checked the convergent validity by looking at AVE values (average variance extracted). These numbers should be above 0.50 (Hair Jr et al., 2013, p. 110). After that, we establish discriminant validity by checking that all the outer loadings on reflective indicators was higher on the constructs it was measuring than on all the other constructs (Hair Jr et al., 2013, p. 105). The rule is that no other cross loadings should be less than 0.2 below the values in the construct the indicator is meant to measure. Also, we checked that the square root of the average variance extracted of each construct is higher than the other correlation with any other first-order constructs. This is done by using Fornell-Larcker criterion (Fornell & Larcker, 1981, pp. 111,112; Hair Jr et al., 2013; Rai, Patnayakuni & Seth, 2006) and recalculating the output from SmartPLS. Recently, it has been suggested that there are better solutions for establishing discriminant validity. Henseler, Ringle and Sarstedt (2015) suggests that Heterotrait-monotrait ratio (HTMT) is better than the abovementioned methods and we therefore, in addition to previous methods, also include this. Now, there are different opinions and unclarity of what threshold to use in order to establish if there are discriminant validity or not. Some authors suggest the threshold should be a value of 0.85 (Clark & Watson, 1995; Kline, 2011) while others propose that a value of 0.90 is better (Gold & Arvind Malhotra, 2001; Teo, Srivastava & Jiang, 2008) If the HTMT's values are under the threshold it suggests that there is discriminant validity.

Analysing the structural model (Inner model)

Before we could analyse and evaluate the inner model, we had to perform our measurement in accordance with the so-called "two-step" (or two-stage) approach, which is suitable for *formative-formative* or *reflective-formative* constructs (Hair Jr et al., 2013, p. 233). The "two-step" approach opposes to the "repeated indicator" approach which is used for reflected second-order constructs. If we were to use the "repeated indicator" approach in our model, the second-order construct would be fully explained by the first-order construct ($R^2 \approx 1.0$), and thereby swamping out other potential effects (Lowry & Gaskin, 2014, p. 135). In our model, that would result in corrupt values for much of the inner model. The "two-step" approach prevented this from happening by performing separate assessments of the measurement model (outer model) and the structural model (inner model) and thus became our choice of approach (Gaskin, 2012; Hair, Ringle & Sarstedt, 2011). In SmartPLS, this was done by running a PLS calculation and extracting the latent variable data (values) from the second and third-order constructs and manually inserting them into the smartPLS' datafile (surveyanswers-coded.csv). Then create a new model with only the four variables (*big data analytics capabilities*, *dynamic capabilities*, *operational capabilities* and *competitive performance*) and the environmental variables

(Hostility, complexity and dynamism). See figure 13 in Appendix 4 for the research model after the two-step reduction.

To assess reliability and validity of the inner model (structural model) we first assessed collinearity by checking the Variance Inflation Factor (VIF) (Hair Jr et al., 2013, pp. 126, 170). Values above 5 indicates high correlation ant therefore these values should be below this limit. Then we looked at the path coefficient and the associated t-values and p-values. This was again extracted through SmartPLS through running PLS calculation and bootstrapping (two tailed test). Then we checked coefficient of determination (R^2) (Hair Jr et al., 2013, pp. 174-177). This was done by running PLS calculation on SmartPLS and extracting these values from the dependent variables. The values range from zero to one and higher levels indicate higher predictive accuracy. The R^2 value is though interpreted differently in depending on the belonging research discipline. As a rule of thumb R^2 values of 0.75 is considered substantial, values of 0.50 is moderate and 0.25 is weak (Hair et al., 2011).

After that we checked the predictive relevance (Q2). This is a measure of how well the path model can predict the originally observed values (Hair Jr et al., 2013, p. 183). This is done by running Blindfolding in SmartPLS and extracting the values from the dependent variables. A value larger than zero indicate the path model's predictive relevance (Hair Jr et al., 2013, p. 178). Now the next test, the effect size (f2) and (q2) of path coefficients is a little trickier to perform. Effect size (f2) is related to endogenous constructs and the effect an omitted construct might have on the constructs R2 value (Hair Jr et al., 2013, p. 177). We tested what the effect would be on *competitive performance* when including and excluding the *dynamic capabilities* construct. The same were done by including and excluding the *operational capabilities* construct. To calculate this the formula used is:

$$f^2 = \frac{R^2_{\text{included}} - R^2_{\text{excluded}}}{1 - R^2_{\text{included}}},$$

The relative impact of predictive relevance can be compared by means of the measure to the q2 effect size (Hair Jr et al., 2013, p. 183). Similar to f2 effect size, this also has to be calculated outside SmartPLS. The values are sorted as follows; 0.02 = small, 0.15 = medium and 0.35 = large. The way this is calculated is:

$$q^2 = \frac{Q^2_{\text{included}} - Q^2_{\text{excluded}}}{1 - Q^2_{\text{included}}}.$$

Finally, the plan consists of a way to test our hypotheses. Since all reliability and validity is conducted the only thing that remains is checking the path coefficients and the significance and relevance of the weights. A summary of the inner model tests can be viewed in table 13 and Appendix 5 include a description of the used tests and sources.

Table 13: Action plan, inner model

Inner model	Structural model	Collinearity Assessment	VIF (Variance Inflation Factor)
		Path Coefficients	Weights
		Coefficient of determination	R2
		Predictive relevance	Q2
		Effect size of path coefficients	Effect size (f2)
			Effect size (q2)
	Hypothesis testing	Path coefficients	Weights
		Significance and relevance of weights	T-values
			P-values

During initial significance testing, all nonparametric "bootstrapping" procedures were based on 500 subsamples, meaning that the PLS-SEM calculation randomly selected 500 subsamples of the total cases. Due to the small number, the results would differ slightly in every recalculation. As for all the final "bootstrapping" procedures, making the basis for the included values in this thesis, the subsample was set to 5000 as recommended by Garson (2016, p. 93) and Hair Jr et al. (2013, p. 156).

Path Coefficient Interpretation

The calculated path coefficients had to undergo a qualitative interpretation in order for us to determine whether the influential effects were small, medium or high. According to (Hair et al., 2011), the different path coefficients in the structural model may be interpreted as standardized beta coefficients from ordinary least squares (OLS) regression. The standardized beta coefficient (or regression coefficient) expresses the average changes in the standard deviation of the dependent variable because of a one-unit change in the standard deviation to the explanatory variable. Under normal circumstances, the beta coefficient will range between –1 and +1. The closer the value is to its extremes, the stronger is the effect (Midtbø, 2013, p. 102).

As for an interpretation of this standardized path coefficient, Kline (2005, p. 122) provide some guidelines that is meant to suite new research areas. Path coefficients values indicates the following:

path coefficient weights < 0.10 indicates a small effect

path coefficient weights around 0.30 indicates a medium effect

path coefficient weights ≥ 0.50 indicates a large effect

Kline (2005, p. 122) further points out that these thresholds should not be interpreted to the extent where a weight of 0.49 and 0.51 is treated differently.

PLS-Multigroup Analysis (PLS-MGA)

As we were aware of the potential lag effects on realizing business value from *big data analytics* initiatives, we wanted to see if such an effect was present in our data. To analyse this, we performed a PLS-Multigroup Analysis (PLS-MGA) which could be used to confirm whether population based parameters would affect the structural models path coefficients (Henseler, Ringle & Sinkovics, 2009, p. 308).

Prior to running the analysis, we had to determine the two groups we wanted to compare. As (Schryen, 2013) contended, it may take up to years to before IS investments gives realized business benefits. We therefore considered it appropriate to split our data between companies that had used *big data analytics* for 0-2 years and those who had used it for 2+ years. This division were based on the background question (BG1) that measured how long the company had used *big data analytics*. This gave us 45 data records in the first data group (0-2 years) and 62 data records in the second data groups (2+ years).

In the PLS-MGA significance test, the significance level was set to 0.05. If the p-value of the path coefficient differences were lower than 0.05 or higher than 0.95, it would indicate that there is a significant difference between the two segmented groups (Garson, 2016, p. 180). If there were to be significant differences between the two tested groups, we would first have to evaluate how big the differences were, to then consider whether parts of the data should be filtered out of the analysis to not damage our results.

4.5 Research Ethics

For this study, we obtained a lot of information from a good number of informants either through interviews or through filling out questionnaires. The information companies share with us is largely mediated by benevolence. To safeguard participants' values, it is important that we as researchers act in a way that is "correct" and ethical. Violation of ethical guidelines would not only damage our reputation as researchers, but weaken the trust of the whole research community we belong to at the University of Agder. Rebuilding such trust may take long time and could thus prevent other researchers from doing their work (Israel & Hay, 2006, pp. 3, 4). With this in mind, it has been highly prioritized to act morally trustworthy as scientists.

Throughout the study, we made sure that all involved participants had made an informed consent to participate. An informed consent implies that participants first understand the purpose of the research and then voluntarily agree to participate (Israel & Hay, 2006, p. 61).

In the initial qualitative phase of the study, we informed all participants about the purpose of the study and assured them that the information would be processed anonymously. During this phase, we also used tape recorders to record the interview, this was also done with the approval from the participants. We also informed them that they were fully entitled to end the interview at any time if it would be desirable. Following the interviews, we sent out transcripts to all the participants where they got the opportunity to correct statements or highlight expressions that they did not want to be quoted on.

Also in the quantitative part, it was important for us that the respondents made an informed consent. All the contacted companies were informed about the nature of the research and that all collected data would be handled anonymously, both at the individual level and company level.

With respect for other researchers and their works, we have done our outmost to avoid plagiarism. In this context, we have been very careful to credit other researchers' work by referencing according to the APA 6^{th} standard (Oates, 2006, p. 61).

5. Analysis and results

In this chapter, we present the results of our analysis efforts. These results are presented in as plainly, simply and transparently a matter as possible while still providing a basic understanding of what the data show. First, we present the results from the exploratory case study. Then, we present the results of the quantitative survey results. These include demographic data, reliability and validity and hypotheses testing. Finally, we provide a summary of this chapter.

5.1 Exploratory case study results

Many interviewees said that one of the biggest challenges associated with *big data* is having access to data. The challenges often arose because the different systems and the data were not always compatible between multiple systems or companies. Data that was used were, for instance, their own company data, purchased data, customer data, merged data, aggregated data.

The technologies used were sometimes well-known business intelligence and database tools. Several explained the importance of new solutions based on open source technology (e.g. Hadoop). Furthermore, services from Amazon (Infrastructure as a service, web services) or Adobe analytics were purchased to support the organizations' *big data* department and solutions. Programs developed by the organizations themselves were also used, while this was typical for more mature *big data* projects.

Several participants talked about the importance of investing resources like money and time in *big data* projects. There were several reasons for this. One reason was that *big data* solutions are viewed as an experimental process. Other reasons were that the phenomenon is relatively new and the solutions are constantly evolving, something that may require high focus on self-development and coursing.

When it came to an organization's technological knowledge, several talked about the knowledge needed to handle the enormous amount of data. Several talked about the need to to know about programming languages like Java, R, Python and others. In addition, knowledge about open source programs such as Hadoop, Cassandra, MongoDB and others must be provided. Several commented that the ability to learn was important. This is because new solutions and systems are constantly presented and implemented in the organization's *big data* departments; especially since it's a fast-moving field. In addition, existing technology solutions such as SQL and data warehouses provide important knowledge about the *big data* world. Several said that knowledge is scarce and hard to retain since competition also searches for the same competency in employees.

Several interviewees explained that top management support was sometimes a challenge. It is not always understood that *big data* initiatives are still experimental in business and there might be resistance among different managers when it comes to implementing *big data* solutions in existing business processes. Trusting their own intuition might overshadow trusting results from *big data analytics*. Some said that to promote a data-driven culture demands high quality of data and tools.

big data can also lead to changes and *big data analytics* can be a driving force that makes an organization agile or increases agile capabilities. Three of the organizations did use *big data analytics* in their day to day processes (or operational processes). They were all in the same industry (media industry). The other three did use *big data* in a more experimental way where it was used as a supplement in strategic decisions. All interviewees said that using *big data analytics* solutions increased their competitive advantage, at least for now. In addition, those organizations that used *big data analytics* in operational processes said that their industry demanded that they use those products. If they didn't, they would lose marked shares to competitors that use *big data analytics*. Also, industries might have suppliers with tailored *big data analytics* solutions that these organizations might buy. For some of the interviewees, the competitive environment demanded that these or similar solutions were used to uphold their competitive status.

Regarding the value of *big data*, one interviewee said it nicely: "We want to substantiate everything we offer (customers) with data (proof). That is the value we try to achieve".

5.2 Survey Analysis and Results

In this section, we will present the outcome of our analysis efforts which was centered around finding ways to establish if our hypotheses could be supported or not.

5.2.1 Demographic data

Of the 557 contacted companies, 525 surveys were sent to CIOs, CEOs, managers and head of *big data* departments (one survey per company). 134 participants completed the survey and 107 of those answered that they use *big data* solutions.

Our selection of participants consists of a wide range of organizations in different industries. There is a slight predominance of industries such as *Media*, *Consumer Goods*, *ICT and Telecommunications*, *Technology* and *Bank and Financials*. The size of the organizations is in most cases defined as large, i.e. among the largest in the Nordic countries. Most are also well established organizations with many years of operation. Our first question in the survey asked if they use *big data* solutions in their organization (all that answered no were excluded from further analysis and removed from the sample). Then followed a question asking how long they had used such solutions. The answers show that there is a large amount of diversification among organization's user experience on *big data* solutions. Surprisingly many have been running *big data* solutions for over 4 years. For an overview of the 107 participants and the associated demographic questions and measurements, see table 14.

Table 14: Demographic data

Dimension	Population (n)	Frequency (%)
Industry		
Media	15	14.02 %
Consumer Goods	14	13.08 %
ICT and Telecommunications	14	13.08 %
Industrials (Construction & industrial goods)	11	10.28 %
Technology	10	9.35 %
Bank and financials	8	7.48 %
Transport	8	7.48 %
Oil & Gas	7	6.54 %
Utilities	7	6.54 %
Consumer Services	4	3.74 %
Basic Materials (Chemicals, paper, industrial metals & mining)	2	1.87 %
Shipping	2	1.87 %
Consulting Services	1	0.93 %
Health Care	1	0.93 %
Other...	3	2.80 %
Big Data Analytics Experience		
< 1 year	16	14.95 %
1 – 2 years	29	27.10 %
2 – 3 years	16	14.95 %
3 – 4 years	12	11.21 %
4+ years	34	31.78 %
Age of Company		
< 1 year	0	0 %
1 – 4 years	7	6.54 %
5 – 9 years	7	6.54 %
10 – 49 years	43	40.19 %
50+ years	50	46.73 %
Size of Company		
0 – 9 employees	0	0 %
10 – 49 employees	13	12.15 %
50 – 249 employees	17	15.89 %
250+ employees	77	71.96 %

5.2.2 Reliability and validity

The research model used is based on a deductive approach. The variables and indicators are mostly based on earlier research and theory found via the literature review. At the same time, we should mention that some of the research, the part dealing with the exploratory case study, is of a more inductive approach. The knowledge that came from this part of the research is also an influencing factor of both model development and selection of indicators that evaluated the model and indicators that were used. Our supervisors acted as advisors so that our construct and measurement further could be refined. The final model consists of different parts that has been researched and tested via peer reviewed articles from quality journals. The challenge, however, was to put together the model so that research questions could be answered and the hypotheses could be measured. We knew the construction of the different variables was of good quality (performed by earlier researchers with success) but the composition of the model had not been validated by any previous researchers. Our indicators are mostly reflective but some of the indicators in our *big data analytics capabilities* first-order construct are formative.

In the next sections, analytical results are presented. View Appendix 5 for an overview of the tests performed, this Appendix includes briefly explanations and sources.

Evaluation of the measurement models (Outer model)

All indicators, except the three used to measure the first-order latent variable *Seize* (*dynamic capabilities*), were collected from peer reviewed articles published in quality journals. The items used to measure *Seize* were carefully built and reviewed based on literature from experts of *dynamic capabilities* and evaluated by our supervisors before they were accepted as good measurement indicators.

Formative measures

We used our tool (SmartPLS) to calculate the path coefficients (weights). To further establish the validity and reliability of the outer model we calculated the t-values of all the formative indicators as a two-tailed (two sided) test (Hair Jr et al., 2013, p. 172; UCLA, 2017). P-values over 0.05 are included in the tables as additional info to show the relevance of the weights. Then we assessed Variance Inflation Factor (VIF). Values should be below 3.3 (Petter et al., 2007). Adequacy coefficient (R^2_a) were calculated and these values should be above 0.50 (Edwards, 2001). See table 15 and 16 for an overview of the results from the first, second and third-order of the multi-ordered constructs.

Table 15: Formative indicators value

Latent variable	Indicator	Weight	T-value	P	VIF	R^2_a
Basic resources	BR1	0.540	2.330	p<0.05	1.648	0.81
	BR2	0.568	2.548	p<0.05	1.648	
Data	D1	0.365	1.768	**p<0.1**	1.064	0.54
	D2	0.423	1.989	p<0.05	1.318	
	D3	0.556	3.014	p<0.01	1.336	
Technology	T1	0.011	0.032	**n.s.**	2.008	0.60
	T2	0.649	2.963	p<0.01	1.232	
	T3	0.530	1.760	**p<0.1**	2.062	

Table 16: Formative measurement second and third-order construct

Construct	Measures	Weight	T-value	P	VIF	R^2_a
Tangible	Data	0.525	3.337	p<0.001	1.483	0.61
	Basic resources	0.625	5.041	p<0.001	1.098	
	Technology	0.146	0.884	**p<0.4**	1.419	
Human skills	Managerial skills	0.714	6.379	p<0.001	1.466	0.77
	Technical skills	0.405	3.031	p<0.01	1.466	
Intangible	Data-driven culture	0.538	3.361	p<0.01	1.160	0.74
	Organizational learning	0.666	4.480	p<0.001	1.160	
BDAC	Tangible	0.650	4.333	p<0.001	2.070	0.56
	Human skills	0.173	0.977	**p<0.4**	1.985	
	Intangible	0.312	2.009	p<0.05	1.650	
Dynamic capabilities	Sense	0.424	2.565	p<0.001	1.622	0.70
	Seize	0.179	1.253	**p<0.3**	1.503	
	Transform	0.575	4.009	p<0.001	1.635	
Operational capabilities	Marketing capabiliy	0.348	1.797	**p<0.1**	1.274	0.71
	Technological capability	0.790	5.143	p<0.001	1.274	

We see that there are some insignificant values between the indicators and the first-order latent variables (D2, T1 and T3) and between technology and tangibles in the second-order and between human skills and *big data analytics capabilities* in the third-order. Although some might suggest removing them we urge that they are very important for the constructs and therefore, we chose to keep them in our model. This is supported by Cenfetelli and Bassellier (2009) , who explains that in models with formative constructs and many indicators, it is likely that there are several that may be insignificant and that unlike reflective constructs, they can be retained as long as the researchers can justify the contribution of it.

Reflective measures

We measured composite reliability, Cronbach's alpha and indicator reliability. Composite reliability and Cronbach's alpha should have values above 0.708 and each reflected indicator should have a loading above 0.708 (Hair Jr et al., 2013, pp. 109,115). We removed three indicators that were below the threshold (loadings < 0.708). These were MC2, CP1 and CP6. See table 17 for an overview of the results.

Table 17: Composite reliability, Cronbach's alpha and indicator reliability

Latent variable	Indicator	Loadings	Cronbach's alpha	Composite reliability
Technology skills	TS1	0.945	0.889	0.947
	TS2	0.952		
Managerial skills	MS1	0.843	0.851	0.910
	MS2	0.916		
	MS3	0.872		
Organizational learning	OL1	0.881	0.727	0.880
	OL2	0.891		
Data-driven culture	DD1	0.828	0.744	0.854
	DD2	0.791		
	DD3	0.820		
Sense	DS1	0.714	0.734	0.850
	DS2	0.848		
	DS3	0.859		
Sieze	DZ1	0.905	0.862	0.916
	DZ2	0.938		
	DZ3	0.810		
Transform	DT1	0.862	0.895	0.935
	DT2	0.927		
	DT3	0.936		
Marketing capabilities	MC1	0.759	0.662	0.815
	MC3	0.808		
	MC4	0.746		
Technological capabilities	TC1	0.751	0.775	0.870
	TC2	0.901		
	TC3	0.837		
Environment complexity	EC1	0.711	0.653	0.829
	EC2	0.960		
Environment hostility	EH1	0.862	0.738	0.883
	EH2	0.915		
Environment dynamism	ED1	0.796	0.756	0.859
	ED2	0.887		
	ED3	0.769		
Competitive performance	CP2	0.735	0.854	0.895
	CP3	0.854		
	CP4	0.860		
	CP5	0.790		
	CP7	0.728		

Convergent validity is assessed by looking at AVE values (average variance extracted). These numbers should be above 0.50 (Hair Jr et al., 2013, p. 110) and are included in table 19.

We established discriminant validity by creating a cross loading overview and checked that the indicators measured what they should measure. Se table 18 for an overview of the cross loadings.

Table 18: Cross loadings

Items	Basic resources	Data	Technology	Managerial skills	Technology skills	Data-driven cult.	Org. Learning	Sense	Sieze	Transform	Marketing cap.	Technological cap.	Competitive perf.	E. Complexity	E. Hostility	E. Dynamism
BR1	**0.897**	0.269	0.225	0.363	0.516	0.247	0.350	0.359	0.320	0.320	0.321	0.412	0.311	0.116	0.059	0.072
BR2	**0.907**	0.259	0.158	0.377	0.312	0.336	0.297	0.313	0.254	0.354	0.298	0.387	0.414	0.154	0.190	0.161
D1	0.174	**0.572**	0.485	0.294	0.320	0.176	0.298	0.230	0.232	0.111	0.161	0.302	0.098	0.040	0.274	0.251
D2	0.238	**0.762**	0.250	0.396	0.442	0.279	0.313	0.188	0.329	0.342	0.108	0.191	0.397	0.144	0.209	0.135
D3	0.231	**0.843**	0.463	0.529	0.343	0.193	0.522	0.273	0.265	0.331	0.050	0.258	0.440	0.125	0.158	0.388
T1	0.235	0.340	**0.632**	0.229	0.223	0.004	0.237	0.112	0.215	0.119	0.055	0.201	0.013	0.055	0.115	0.226
T2	0.179	0.461	**0.872**	0.318	0.297	0.265	0.337	0.313	0.188	0.184	0.150	0.353	0.126	-0.056	0.149	0.103
T3	0.175	0.449	**0.806**	0.308	0.220	0.182	0.341	0.233	0.173	0.199	0.049	0.283	0.132	0.105	0.092	0.234
MS1	0.316	0.484	0.342	**0.845**	0.412	0.305	0.519	0.336	0.363	0.399	0.197	0.203	0.386	0.190	0.262	0.247
MS2	0.398	0.524	0.342	**0.916**	0.564	0.270	0.446	0.186	0.359	0.369	0.227	0.337	0.346	0.127	0.219	0.132
MS3	0.364	0.490	0.297	**0.872**	0.503	0.326	0.403	0.271	0.427	0.434	0.208	0.385	0.328	0.046	0.105	0.113
TS1	0.406	0.457	0.307	0.517	**0.945**	0.342	0.360	0.205	0.318	0.321	0.143	0.296	0.240	0.010	0.083	0.086
TS2	0.458	0.480	0.286	0.552	**0.952**	0.325	0.394	0.180	0.233	0.416	0.231	0.397	0.236	0.007	0.145	0.087
DD1	0.165	0.266	0.159	0.267	0.210	**0.828**	0.284	0.268	0.328	0.353	0.232	0.372	0.229	0.002	0.070	0.071
DD2	0.264	0.138	0.152	0.166	0.227	**0.791**	0.299	0.345	0.205	0.232	0.304	0.249	0.229	-0.044	0.196	0.039
DD3	0.357	0.293	0.331	0.386	0.408	**0.820**	0.323	0.303	0.183	0.295	0.214	0.355	0.242	0.049	0.170	0.180
OL1	0.220	0.529	0.405	0.493	0.361	0.312	**0.881**	0.277	0.414	0.293	0.155	0.386	0.192	-0.069	0.029	0.196
OL2	0.410	0.416	0.309	0.426	0.344	0.347	**0.891**	0.299	0.299	0.338	0.161	0.296	0.242	0.152	-0.012	0.184
DS1	0.231	0.119	0.256	0.208	0.081	0.230	0.262	**0.714**	0.437	0.411	0.315	0.377	0.379	0.111	0.103	0.162
DS2	0.333	0.316	0.242	0.272	0.195	0.326	0.276	**0.848**	0.357	0.550	0.465	0.390	0.446	0.142	0.336	0.281
DS3	0.330	0.311	0.299	0.240	0.203	0.344	0.253	**0.859**	0.452	0.403	0.627	0.421	0.494	0.270	0.263	0.165
DZ1	0.220	0.347	0.184	0.362	0.262	0.212	0.357	0.452	**0.905**	0.463	0.208	0.439	0.299	0.109	0.106	0.133
DZ2	0.296	0.382	0.258	0.420	0.286	0.339	0.438	0.511	**0.938**	0.517	0.265	0.451	0.362	0.129	0.179	0.211
DZ3	0.335	0.244	0.115	0.376	0.214	0.213	0.251	0.378	**0.810**	0.374	0.207	0.441	0.318	0.096	0.151	0.166
DT1	0.310	0.263	0.214	0.441	0.298	0.270	0.315	0.507	0.479	**0.862**	0.338	0.413	0.440	0.128	0.124	0.144
DT2	0.365	0.379	0.231	0.394	0.343	0.365	0.341	0.533	0.472	**0.927**	0.315	0.418	0.580	0.138	0.166	0.270
DT3	0.342	0.359	0.172	0.411	0.421	0.349	0.315	0.500	0.454	**0.936**	0.301	0.356	0.534	0.106	0.145	0.229
MC1	0.396	0.111	0.034	0.252	0.187	0.348	0.110	0.511	0.300	0.366	**0.759**	0.433	0.363	0.287	0.257	0.072
MC3	0.256	0.096	0.036	0.170	0.161	0.256	0.190	0.459	0.143	0.278	**0.808**	0.362	0.311	0.250	0.141	0.160
MC4	0.112	0.098	0.245	0.121	0.103	0.073	0.111	0.375	0.141	0.138	**0.746**	0.260	0.219	0.094	0.227	0.060
TC1	0.235	0.315	0.282	0.268	0.206	0.339	0.217	0.431	0.552	0.291	0.289	**0.751**	0.341	-0.099	0.032	-0.046
TC2	0.437	0.253	0.335	0.273	0.356	0.390	0.407	0.370	0.425	0.401	0.396	**0.901**	0.298	-0.039	-0.019	0.042
TC3	0.414	0.275	0.331	0.336	0.340	0.279	0.321	0.425	0.293	0.384	0.460	**0.837**	0.318	0.062	0.205	0.120
CP2	0.406	0.416	0.224	0.425	0.320	0.184	0.380	0.332	0.298	0.346	0.348	0.384	**0.735**	0.227	0.161	0.184
CP3	0.267	0.334	0.125	0.302	0.195	0.115	0.205	0.385	0.271	0.384	0.202	0.239	**0.854**	0.181	0.036	0.151
CP4	0.283	0.335	0.092	0.292	0.182	0.220	0.150	0.425	0.249	0.448	0.307	0.215	**0.860**	0.259	0.291	0.175
CP5	0.285	0.407	0.134	0.333	0.204	0.185	0.168	0.509	0.411	0.614	0.194	0.363	**0.790**	0.093	0.227	0.256
CP7	0.356	0.283	0.043	0.249	0.109	0.234	0.099	0.477	0.219	0.433	0.489	0.296	**0.728**	0.296	0.323	0.226
EC1	0.204	0.080	0.012	0.099	0.062	0.040	0.077	0.111	0.000	0.005	0.340	0.050	0.114	**0.711**	0.214	0.201
EC2	0.106	0.148	0.020	0.130	-0.014	-0.009	0.031	0.226	0.157	0.167	0.216	-0.052	0.285	**0.960**	0.323	0.517
EH2	0.090	0.293	0.158	0.183	0.146	0.110	0.062	0.312	0.130	0.154	0.288	0.033	0.214	0.239	**0.862**	0.309
EH3	0.152	0.210	0.110	0.210	0.078	0.198	-0.035	0.229	0.161	0.134	0.202	0.119	0.268	0.335	**0.915**	0.345
ED1	0.120	0.358	0.143	0.158	0.116	0.061	0.220	0.176	0.207	0.165	0.052	0.019	0.179	0.333	0.236	**0.796**
ED2	0.129	0.291	0.176	0.199	0.139	0.224	0.210	0.294	0.162	0.289	0.188	0.101	0.253	0.454	0.331	**0.887**
ED3	0.065	0.259	0.153	0.083	-0.051	-0.028	0.090	0.126	0.111	0.101	0.045	-0.013	0.184	0.378	0.335	**0.769**

We checked the Fornell-Larcker criterion (Fornell & Larcker, 1981, pp. 111,112; Hair Jr et al., 2013; Rai et al., 2006) method and recalculating the output from SmartPLS. See table 19 for the results using this method.

Table 19: Fornell-Larcker criterion

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Basic resources	**n/a**															
2. Competitive performance	0.403	**0.892**														
3. Data	0.293	0.448	**n/a**													
4. Data-driven culture	0.325	0.241	0.290	**0.902**												
5. Environment Complexity	0.150	0.266	0.145	0.006	**0.919**											
6. Environment Dynamism	0.130	0.255	0.365	0.123	0.480	**0.905**										
7. Environment Hostility	0.140	0.273	0.277	0.178	0.328	0.369	**0.943**									
8. Managerial skills	0.410	0.402	0.569	0.341	0.137	0.185	0.222	**0.937**								
9. Marketing capabilities	0.343	0.392	0.132	0.305	0.283	0.128	0.269	0.240	**0.879**							
10. Organizational learning	0.358	0.245	0.532	0.372	0.049	0.215	0.009	0.518	0.178	**0.941**						
11. Sense	0.372	0.545	0.316	0.374	0.217	0.254	0.298	0.298	0.587	0.325	**0.900**					
12. Sieze	0.318	0.369	0.371	0.293	0.127	0.194	0.165	0.436	0.258	0.401	0.509	**0.941**				
13. Technological capabilities	0.442	0.381	0.335	0.403	-0.025	0.052	0.091	0.353	0.464	0.384	0.489	0.499	**0.912**			
14. Technology	0.211	0.152	0.541	0.268	0.020	0.193	0.147	0.372	0.124	0.402	0.327	0.216	0.381	**n/a**		
15. Technology skills	0.456	0.251	0.495	0.351	0.009	0.092	0.121	0.564	0.199	0.397	0.203	0.289	0.367	0.312	**0.974**	
16. Transform	0.374	0.573	0.369	0.363	0.136	0.239	0.160	0.455	0.348	0.356	0.565	0.514	0.435	0.226	0.390	**0.953**
Mean	4.54	4.67	5.06	4.68	4.83	4.46	4.97	5.03	4.79	5.26	4.72	4.77	4.57	4.64	4.85	3.79
Standard deviation	1.46	1.40	1.62	1.39	1,26	1.25	1.26	1.43	1.28	1.43	2.09	1.36	1.40	1.69	1.81	1.95
AVE	n/a	0.631	n/a	0.661	0.771	0.595	0.785	0.655	0.785	0.692	n/a	0.900	0.827	0.713	0.671	0.790
Cronbach's alpha	n/a	0.854	n/a	0.744	0.851	**0.662**	0.727	0.734	0.862	0.775	n/a	0.889	0.895	**0.653**	0.756	0.738
Composite reliability	n/a	0.895	n/a	0.854	0.910	0.815	0.880	0.850	0.916	0.870	n/a	0.947	0.935	0.829	0.859	0.883

We also checked that all the values in Heterotrait-monotrait ratio (HTMT) (Henseler et al., 2015) passed as good. There are different opinions and unclarity of what threshold to use to establish if there are discriminant validity or not. Some authors suggest the threshold should be a value of 0.85 (Clark & Watson, 1995; Kline, 2011) while others propose that a value of 0.90 is better (Gold & Arvind Malhotra, 2001; Teo et al., 2008). All our values are below 0.85 and verified as acceptable. To view the results from the HTMT see table 20.

Table 20: Heterotrait-monotrait ratio

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Competitive performance													
2. Data-driven culture	0.305												
3. Environment Complexity	0.355	0.081											
4. Environment Dynamism	0.305	0.205	0.584										
5. Environment Hostility	0.326	0.235	0.442	0.489									
6. Managerial skills	0.474	0.423	0.179	0.238	0.280								
7. Marketing capabilities	0.507	0.420	0.477	0.161	0.394	0.312							
8. Organizational learning	0.320	0.505	0.183	0.286	0.075	0.662	0.255						
9. Sense	0.676	0.504	0.311	0.322	0.401	0.379	0.824	0.447					
10. Sieze	0.425	0.361	0.131	0.240	0.204	0.510	0.331	0.500	0.642				
11. Technological capabilities	0.467	0.530	0.135	0.107	0.173	0.432	0.630	0.507	0.655	0.627			
12. Technology skills	0.292	0.427	0.058	0.159	0.154	0.645	0.252	0.494	0.246	0.330	0.434		
13. Transform	0.639	0.440	0.160	0.271	0.199	0.525	0.440	0.441	0.695	0.583	0.520	0.435	

Evaluation of the structural model (Inner model)

To assess reliability and validity of the Structural model (inner model) we looked at the VIF (variance inflation factor). This is calculated by SmartPLS. Further, we checked the path coefficient and the associated t-values and p-values. Coefficient of determination values were extracted from SmartPLS by using PLS calculation and Predictive relevance (Q2) by using the function "Blindfolding". See table 21 for an overview.

Table 21: Simplified inner model

Paths	Weight	T-value	P-value	Coefficient of determination (R2)	Predictive relevance (Q2)	VIF
BDAC --> DC	0.585	8.794	p<0.001	0.336 (DC)	0.211 (DC)	1.000
BDAC --> OC	0.546	4.187	p<0.001	0.291 (OC)	0.163 (OC)	1.000
DC --> CP	0.450	7.239	p<0.001	0.430 (CP)	0.254 (CP)	2.228
OC --> CP	0.071	0.629	**n.s.**			1.917
BDAC --> CP	0.146	1.203	**p<0.4**			1.948

The next step was to calculate the effect size (f2) and effect size (q2). We focused on the latent variable representing *competitive performance*. See table 22 for the results.

Table 22: Effect size (f2) and (q2) on competitive performance (CP)

		CP (Included)	
Excluded	Path coefficient	f2 effect size	q2 effect size
DC	0.450	0.158	0.070
OC	0.146	-0.019	-0.003

Even if there are no hypotheses in our model associated directly with the moderating environmental factors, they might help explaining the relationships presented in the hypotheses. The moderating factors and associated data can be seen in table 23.

Table 23: Moderating effect on DC to CP and OC to CP

Moderator	Path	Weight	T-value	P	VIF
E. hostility	DC --> CP	-0.190	1.371	**p<0.2**	2.547
	OC --> CP	0.140	0.938	**p<0.4**	2.728
E. complexity	DC --> CP	-0.014	0.135	-	2.116
	OC --> CP	0.075	0.652	-	1.999
E.dynamism	DC --> CP	-0.003	0.024	-	2.994
	OC --> CP	-0.207	1.545	**p<0.2**	2.730

5.2.4 Time Lag

In this section, we will present the PLS-MGA results which identified whether there were significant differences between companies that recently have adopted *big data analytics* (0-2 years) and those who have had it for a longer time (2+ years). The path coefficient differences between the two data segmentation groups listed in the second column (table 24) shows the differential weights spanning from 0.063 to 0.247. This expresses a small positive segment difference onto companies who had used it for more than 2 years. However, as none of the measured p-values were lower than 0.05 or larger than 0.95, these differences were not significant. On the basis of these nonsignificant values, we decided to use the full sample pool of 107 respondents.

Table 24: Time lag, measured differences

Path	Path Coefficient-diff	P-value
BDAC --> DC	0.063	0.310
BDAC --> OC	0.167	0.863
DC --> CP	0.205	0.750
OC --> CP	0.247	0.829

5.2.3 Testing hypotheses

After establishing the validity and reliability of the research model, hypotheses were evaluated. Now, the four hypotheses in our model were constructed to see if there are correlation between the following elements: *big data analytics capabilities*, *dynamic capabilities*, *operational capabilities* and *competitive performance*. According to (Hair et al., 2011) path coefficient weights < 0.10 indicates a small effect, around 0.30 indicates a medium effect and ≥ 0.50 indicates a large effect. Each hypothesis and a repetition of important supporting factors will be presented in the following sections. Our research model with additional measures are shown in figure 11. Table 25 shows a summary of the hypotheses and the supporting data.

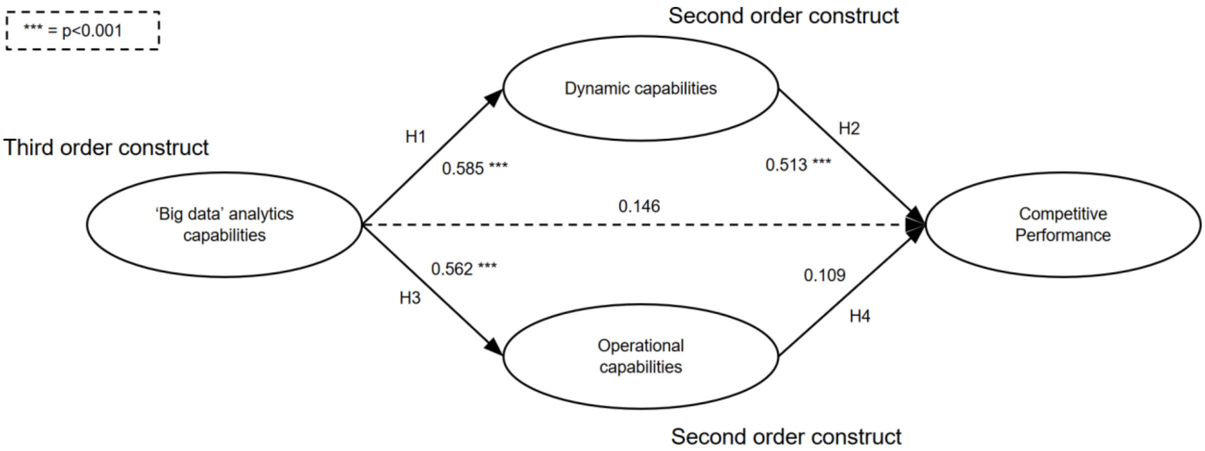


Figure 11: Research model, weights and p-values

Hypothesis 1: *"There is a positive correlation between big data analytics capabilities and dynamic capabilities"*

This hypothesis has a strong weight of 0.585. It is supported with a t-value of 8.794 which significantly over 99.9 percent which equals a p-value of less than 0.001. The reliability and validity was acceptable. Therefore, we can confirm that this hypothesis is **supported**.

Hypothesis 2: *"There is a positive correlation between dynamic capabilities and competitive performance"*

This hypothesis has a strong weight of 0.450. It is supported with a t-value of 7.239 which is significantly over 99.9 percent which equals a p-value of less than 0.001. The reliability and validity was acceptable. Therefore, we can confirm that this hypothesis is **supported**.

Hypothesis 3: *"There is a positive correlation between big data analytics capabilities and operational capabilities"*

This hypothesis has a strong weight of 0.546. It is supported with a t-value of 4.187 which is significantly over 99.9 percent which equals a p-value of less than 0.001. The reliability and validity was acceptable. Therefore, we can confirm that this hypothesis is **supported**.

Hypothesis 4: *"There is a positive correlation between operational capabilities and competitive performance"*

This hypothesis has a weak weight of 0.071. It is supported with a t-value of 0.629 which is under the scientific threshold of 95 percent and equals a p-value of less than 0.05. The reliability and validity was acceptable. Therefore, this hypothesis is **not supported**.

Table 25: Hypothesis and conclusions

Hypothesis	Independent variable	Dependent variable	Weight	T-value	P	Conclusion
H1	BDAC	DC	0.585	8.794	$p<0.001$	Supported
H2	DC	CP	0.450	7.239	$p<0.001$	Supported
H3	BDAC	OC	0.546	4.187	$p<0.001$	Supported
H4	OC	CP	0.071	0.629	**n.s.**	Not supported

6. Discussion

In this chapter, we discuss our findings and compare those with existing literature and earlier studies.

Our study is based on previous research and therefore can be viewed as confirmations to both the measurement of *big data analytics capabilities* and the connection between *Dynamic* and *operational capabilities* and *competitive performance*. The connections between *big data analytics capabilities* and *dynamic capabilities* or *operational capabilities* have not been tested empirically in past research. In addition, to the best of our knowledge, there has not been any similar studies that involves Nordic organizations until now.

We began by summarizing the findings of our research study. Then we discuss our four hypotheses and how the results fit with previous research. Then we discuss the moderating effect before we move on to other findings. After that we discuss the reliability and validity results we acquired from the associated tests. Next, we discuss the implications this research has from a theoretical and practical perspective. We conclude this chapter by discussing limitations and suggestions for future works.

6.1 Summary of research

This research was primarily concerned with explaining how *big data analytics capabilities* investments could lead to value for organizations and the way to measure value was through *competitive performance*. *Big data analytics capabilities* is an interesting and a relative new phenomenon that has received a lot of attention recently. The previous research has been focused on the technical side of what *big data* could provide when it comes to analytical insight. Very little research is done on HOW organizations get value from *big data* or if they add value at all. Would this strengthen their competitive advantage? This is what we wanted our research to determine.

Throughout the literature review, and the exploratory case study, we identified two ways organizations used *big data* solutions. It was through the operational processes as a way to improve day to day processes. This result was the main usage of three of the organizations we interviewed. Earlier literature also supported this. The other way they used *big data* solutions was through more strategic processes. This lead us to change our theoretical lenses from Resource-Based View (RBV), which is more suited to explain what resources an organization has per se, to *dynamic capabilities* which focuses more on an organization's capability to sense, change and learn.

To help explain how *big data*, *dynamic capabilities* and *operational capabilities* react in turbulent environments we implemented three different environmental factors. Earlier research proclaims that these environmental factors provide a moderating effect on the performance outcome of both *dynamic capabilities* and *operational capabilities*. We wanted to verify this when the focus was solely on *big data analytics capabilities*.

To test our hypothesis, we decided to perform a quantitative study on a sample of 107 Nordic organizations. These were analyzed with the aid of Partial Least Squares Structural Equation Modeling (PLS-SEM) by using the tool SmartPLS. The empirical findings will be discussed in the following sections.

6.2 Discussion of the research question and hypotheses

To answer the research question two paths were identified and tested. The path from *big data analytics capabilities* to *competitive performance* through *dynamic capabilities* (H1 and H2) were significant and the path coefficients had moderate to high values. So we got empirical evidence that the *dynamic capabilities* path is a mediator. We can not come to the same conclusion when it comes to the path between *big data analytics capabilities* to *competitive performance* through *operational capabilities* (H3 and H4). Even if we found empirical evidence for a correlation between *big data analytics capabilities* and *operational capabilities*, we did not find any significance on the path between *operational capabilities* and *dynamic capabilities*. The direct path between *big data analytics capabilities* to *competitive performance* was low and insignificant.

The following subsections contains our interpretation of the findings and are structured in accordance to the hypotheses. Our interpretation will be based on the path coefficients that links the hypothesis latent variables and whether this relationship are statistically genuine or down to chance. These results are further used to clarify whether our findings relate to prior literature.

Hypothesis 1: Correlation between big data analytics capabilities and dynamic capabilities

The analysis showed a strong empirical support for the first hypothesis (H1) with a significant ($p < 0.001$) path coefficient weight of 0.585, which indicates a large influential effect. This positive correlating effect between *big data analytics capabilities* and *dynamic capabilities* matches our pre-conceptions. In today's fastmoving business world, information is key to decide the way forward when threats or opportunities occur. Companies are adopting new and sophisticated analytical tools to cope with the abundance of digital information that we today are surrounded by. This finding helps unfolding the black box that wraps the process between investment and success by showing that companies who develop distinctive *big data analytics capabilities* generally has a higher level of *Dynamic Capability* and thus a higher level of evolutionary fitness.

Hypothesis 2: Correlation between dynamic capabilities and competitive performance

The second hypothesis (H2) is also supported with a significant ($p < 0.001$) path coefficient of 0.45, which is seen as a medium influential effect. By sensing, seizing and reconfiguring configuring resources, companies can charge and renew their resource base with competitive resources. This is also consistent with prior literature where most agree that *dynamic capabilities* contribute to *competitive performance* (Drnevich & Kriauciunas, 2011; Pavlou & El Sawy, 2011; Wilden et al., 2013).

Hypothesis 3: Correlation between big data analytics capabilities and operational capabilities

This hypothesis is supported with a path coefficient weight of 0.546 and a p-value below 0.001. This fit well with our assumptions from our case-study, where several organizations expressed this view. There is limited previous research on the association between *big data analytics capabilities* and *operational capabilities.* (Chen et al., 2015) asked for more empirical demonstrations of this association. This result is almost the same as hypothesis one (*big data analytics capabilities* and *dynamic capabilities*). Even if these investments into *big data analytics capabilities* will automatically increase both *dynamic capabilities* and *operational capabilities*, it might be more correct to say that the association depends on the organisation's choice. Do they use *big data analytics* in *operational capabilities*, in *dynamic capabilities* or in both? There is no doubt that a positive correlation between them exists. This empirical result might contribute to more understanding. To invest resources in order to achieve better capabilities, the result still needs further empirical research to establish such a theory. For instance, *dynamic capabilities* might be a source for indirect positive correlation between *big data analytics capabilities* and *operational capabilities* or provide indirect positive correlation between *operational capabilities* and *competitive performance*. This is outside the scope of this research but could further explain the link between *big data analytics capabilities* and *competitive performance*.

Hypothesis 4: Correlation between operational capabilities and competitive performance

While this hypothesis had a path coefficient weight of 0.071 it's p-value was above the significant threshold ($p<0.05$). According to Chin (1998), standardized paths should be at least 0.20 to be considered meaningful for discussion. This and the low p-value leads to into failure of rejecting the null hypothesis and the hypothesis 4 is therefore NOT supported. Several previous articles support the connection between *Operational Capabilities* and *competitive performance* (El Sawy & Pavlou, 2008; Li et al., 2010; Roth & Jackson III, 1995; Teece, 2007; Wilden & Gudergan, 2015; Wu et al., 2010) and even have empirical evidence for their statements (Drnevich & Kriauciunas, 2011; Wilden & Gudergan, 2015). Therefore, this part of the research needs more attention to see what improvements could be made to more accurately measure *operational capabilities* and *competitive performance* in the *big data analytics capabilities* → *operational capabilities* → *competitive performance* association. Another possibility for this result could be that the population, which all uses *big data analytics*, might be in rapidly changing environments. To conclude, other factors could make these answers unpredictable and insignificant and even if our literature review and exploratory case study supported that a correlation exists, we cannot support this based on the analysed results.

6.3 Discussion of the moderating and contextual factors

These moderators in our model did not have any hypotheses but were included to help explaining correlations on hypotheses H2 and H4. None of them were significant. Therefore, we only discuss these shortly.

Environmental hostility

In our model, hostility moderates the correlation between *dynamic capabilities* and *competitive performance* (H2) and *operational capabilities* and *competitive performance* (H4). This moderating effect had a path coefficient weight of –0.190 on H2 and 0.140 on H4. Both measurements are above the threshold of $p<0.05$ and therefore not significant. The earlier empirical tests done in, for instance, Chen et al. (2014) shows that *environmental hostility* can have a negative moderating effect the link between IT-capability and performance. Even if we also got a negative weight on H2 we conclude that it is too much noise in this data and the t-value is too low to be significant.

Environmental complexity

In our model, complexity moderates the correlation between *dynamic capabilities* and *competitive performance* (H2) and *operational capabilities* and *competitive performance* (H4). This moderating effect had a path coefficient weight of –0.014 on H2 and 0.075 on H4. Both measurements are way above the threshold of $p<0.05$ and therefore not significant. Earlier literature (Chen et al., 2014) has found empirical evidence for this as a moderator between capabilities and performance. We did not find any significant values.

Environmental dynamism

In our model, dynamism moderates the correlation between *dynamic capabilities* and *competitive performance* (H2) and *operational capabilities* and *competitive performance* (H4). This moderating effect had a path coefficient weight of –0.003 on H2 and 0.207 on H4. Both measurements are above the threshold of $p<0.05$ and therefore not significant. Like Chen et al. (2014), we could not find any significance in dynamism moderating on the link between capabilities and performance.

6.4 Discussion of the reliability and validity

To measure the reliability and validity of our study, we used several sources as guidance. This included textbooks provided by other courses at the University of Agder and earlier, well-written master theses. In addition to having a basic understanding of how to assess reliability and validity and secure that our research could be a contribution to future research, we needed to learn what peer-reviewed articles in our research area did to assess this. We constructed a plan to assess reliability and validity to our results (see table 12 on page 44 and table 13 on page 47). Now, even if most values were on the good side of the thresholds, there were a few that needed closer attention.

For the formative measurements, four indicators had a t-value that was below the threshold. These values belonged to the third-order construct of *big data analytics capabilities*. This was very difficult to mitigate since each of these indicators are important for the whole measure of the associated variables. The same thing can be said about the measurements in the second-order and in the third-order. The t-values were below the threshold. *Big data analytics capabilities* must consist of human skills, Tangible must consist of Technology. This is empirically measured in earlier research conducted by Gupta and George (2016). Cenfetelli and Bassellier (2009) suggests that in a model with many indicators some will likely be of non-significance value. This is acceptable as long as the researchers can justify the importance of these indicators. Therefore, we include them as a part of our model and analytical results.

Almost all reflective measures achieved only good values. There was one first-order latent variable, Marketing capabilities, that had a Cronbach's alpha value of 0.662. This was just below the most used threshold which is 0.708. With the empirical research performed by Wilden and Gudergan (2015) the construct and indicators should be well designed. Slavec and Drnovsek (2012) and Hair Jr et al. (2013, p. 107) suggests that values greater than 0.60 could be accepted as reliable – in exploratory research. Further, Slavec and Drnovsek (2012) says that alternatively internal consistency can be evaluated also with item-to-total correlations and inter-item correlations. Hair Jr et al. (2013, pp. 101, 102) says that due to limitations with Cronbach's alpha in PLS-SEM, it is better to apply a different measure for internal consistency, which is referred to as composite reliability. Values between 0.70 and 0.90 can be regarded as satisfactory. Marketing capabilities has a value of 0.815. Marketing capabilities is a formative measurement for *operational capabilities* and should not be excluded. When considering all the above mentioned, we feel confident that this latent variable is reliable.

Environmental hostility was also a latent variable that had a lower Cronbach's alpha, 0.653. It also had good values when measured composite reliability. Since it didn't have any significance through the t-value checks, we will not discuss this any further.

All the hypotheses had reliability and validity except hypothesis four (H4). This hypothesis had a t-value of 0.629 which is very low, way under the threshold of 1.984. The coefficient of determination (R2) value was small to medium and the predictive relevance (Q2) was above zero which mean that the model has predictive accuracy and relevance. These values lead us to conclude that H1-H3 were supported and H4 was not supported.

The difficulties in securing that our research model and data will pass these tests prior to collecting the data has been a big challenge since the respondents from the population is limited (an unknown size) and testing on other respondents (that might not know about *big data analytics capabilities* in a business setting) might result in low quality data and therefore not usable in assessing reliability and validity tests. We have laid heavily on earlier research conducted by peer reviewed articles from quality journals to counter this. The challenge comes when the constructs forms a new model and new influences occur and previous research may not have taken those new influences into account in their models.

6.5 Discussion of the other findings

Our research data provided a lot of information and can be a source for further analysis. We shortly present two of our other findings:

1. In our model, we made another connection from *dynamic capabilities* to *operational capabilities* to see if we would get any valid results. From the literature, we studied in this research, there were suggestions that *dynamic capabilities* would affect *competitive performance* indirectly via *operational capabilities* (Pavlou & El Sawy, 2006; Protogerou et al., 2012; Wilden & Gudergan, 2015). Improving *operational capabilities* might strengthen this capability. Changing this might lead to weaker *operational capabilities*, at least temporarily until improvement, learning and effectiveness had time to affect *operational capabilities*. Calculations in SmartPLS showed valid results with a path coefficient weight of 0.477 and a t-value of 4.629 ($p<0.001$). This show that it is a significant positive correlation value.
2. The analysis of the *big data analytics capabilities* construct also shows some interesting values on the indicator T1 which measured whether companies had adopted and explored parallel computing approaches (e.g. Hadoop) to process *big data*. The indicator had a nonsignificant weight of 0.011. However, we see that indicator T1 has the lowest mean value (mean = 4.15) and at the same time, the highest standard deviation value (S.Dev = 2.28) among the indicators used to measure *big data analytics capabilities* (see Appendix 6 for an overview of the measurement items). These values may be due to the fact that many companies still are in the starting phase when it comes to using this type of sophisticated technology. Prior literature highlights that finding people with the right skill set to handle this kind of technology is not an easy task (Chen et al., 2015; Wamba et al., 2015). This also came to the surface during our initial exploratory case study, where one of the interviewees told the following: *"Lack of competencies has to be the most major challenge we face. There is a gap in the skill that is quite remarkable at the moment, and the gap is increasing because there is a lot of new demand for new skills" (Mikalef et al., 2017)*. Regarding these skills, the interviewee further explained that they had to train themselves in these new technologies by stating *"We've done it our self. Thus, own education. Hadoop expertise was very difficult to obtain so we had to actually train people within the company" (Mikalef et al., 2017).*

6.3 Discussion of the research process

In this section, we discuss some of our thoughts regarding the research process conducted in this project.

Exploratory case study

This study provided insights into Norwegian organizations and their *big data* world. We learned that *big data* can provide value for organizations in different ways. Also, operational processes can use *big data analytics* to increase their efficiency and quality. This is also supported by Chen et al. (2015) Then again, *big data analytics* can be used for making organizations agile in addition to supporting agile organizations. While some look at *big data analytics* as an opportunity to get competitive advantage, others view it as necessary to compete against other companies that already uses *big data analytics*. New technologies (e.g. Hadoop and other open source technologies) are important for the *big data* initiatives and knowledge on these are scarce and hard to obtain and keep. This is also supported by (Emani et al., 2015; Gupta & George, 2016; Wamba et al., 2017). Top management responses were is important since they might influence other leaders and decision makers to accept *big data analytics*. This is also supported by (Chen et al., 2015; GalbRaith, 2014; Garmaki et al., 2016; Tallon et al., 2013; Wamba et al., 2017). This study was necessary to get an understanding of how organizations use and think of *big data*. This might not be necessary in other areas of research into information systems where knowledge and previous literature is more mature than the relative fresh *big data* phenomenon.

Literature review

The systematic literature review was divided into two parts. We felt this was a somewhat unconventional solution as we have not seen others do this before. A thorough investigation of *big data analytics capabilities* was required before we concentrated on the road from *big data analytics capabilities* to *competitive performance*. Another reason why we chose this solution was that there exists few earlier researches on *big data analytics capabilities* as opposed to, for instance, *dynamic capabilities* or *competitive performance*. We think, the more mature theories become, the easier it gets to find quality literature on the subject. Later research on the theme we had chosen might not need such an extensive literature review or at least not a divided one.

Data collecting process

This process is the most time-intensive job for the researchers. Some researchers have time to do different tasks while waiting for the respondents to answer the survey as they send out the survey in one bulk (with several reminders at certain dates). In our case this was not true at all. This part was at least as time consuming as the previous planning phase. This might be different in other surveys that have other sample of populations. In our case, we had to try and get as many respondents as possible of the ones that used *big data analytics* solutions. That meant avoiding things like:

1. Sending mail through survey programs that would end up in the spam folder at the recipients.
2. Respondents missing our invitation or perhaps not feeling obliged to participate.
3. Respondents forgetting us.
4. Respondents feeling overwhelmed by the size of the survey or the number of questions – and cancelling (Oates, 2006, p. 228).
5. Respondents not trusting our invitation email and link (Oates, 2006, p. 229).

With these goals, we decided it was too difficult to send out one standardized mail to a group of people since it would be difficult to follow-up who completed, and keep track of who started and who had not started. Also, by testing our tool’s bulk function the mail was defined as spam and in the tests we committed, we saw that these mails were defined as spam an moved to the spam folder in the email clients. Therefore, we sent out an invitation with a unique survey link from our university email addresses, one for each respondent. Also, the inviting text should be easy to read, provide enough information and build confidence. We had two unsuspected hindrances in this phase. There were at least two major spam attacks (fishing attacks) in Norway and Europe (Sarmadawy, 2017; Zachariassen, 2017). This resulted in respondents avoiding emails with links and they would not start our survey. The way to counter this was to talk to the respondents in person (by phone) or via mail. We provided in our invitation mail, our full name, mail and phone number. This resulted in respondents calling us for confirmation or asking us for additional confirmation from our university. All these efforts and the ones described in 4.2.1 (Method for collecting data) were necessary since our population was unclear and limited when it comes to size. This might not be necessary in surveys with bigger population and easier access to representatives.

Some of the respondents answered NO (we do not use *big data analytics* solutions). Why did they still complete the survey? We prepared the text in the inviting mail to explain we are looking for organizations that are using *big data analytics* solutions. In addition, we included our definition of *big data* and *big data analytics* in our surveys first page and asked if they did use *big data* solutions. What we could have done though, is changing the survey so that when answering NO the survey would end and a "thank you for participating" was provided. Then again, several who answered the whole survey did it because they wanted the copy of our report and the benchmark which was something we offered to everyone who completed the survey.

The planning of the analysis process

This part of the research is something that needs improvements. Many of these improvements would not fit into the time frame of this master thesis since they involve processes that takes much longer than the time provided in a master's thesis to conduct. For instance, a trial survey with participants that know about *big data* and could provide close to real answers could produce realistic data of a size that is closer to the real research data That way many of the reliability and validity tests could help refining the survey before initiating the "real" survey. This was performed by (Gupta & George, 2016) with success and this way, it is easier to find faults in the model, data, indicators or other places that is hard to predict theoretically. Also, the collecting phase could last longer as an increased response could give possibilities to be stricter when choosing a limit for, for instance, how long an organization has used *big data*.

6.4 Theoretical and Practical implications

This study has some interesting findings that can be used in both further research and practical use.

We achieved the same results as Gupta and George (2016) did, with some exceptions. We therefore assume our research supports their third-order construct for *big data analytics capabilities*. We think, in retrospect, we would like to improve the formative measurements and perhaps use reflective measurements instead. This is a challenge since their formative questions have good reasoning and should provide good measurements. In our analysis, we had insignificance on some of them though.

Another use for this research is as a guide for further refinement of our model and measures. For instance, our results can provide researchers with deeper understanding of *big data analytics capabilities* and how to improve this third-order construct.

Dynamic capabilities are definitely in correlation with *big data analytics capabilities*. Researchers could now dig deeper into both *big data analytics capabilities* and *dynamic capabilities* and look closer on the building blocks of these capabilities.

We also explain the research phases we developed. Just as we got tips and advice through other articles, master's theses and textbooks, we hope our thesis provides readers with useful tips for conducting, for instance, a similar data collecting phase.

A practical usage is that organizations and especially chief information officers (CIOs) could inspect the *big data analytics capabilities* construct, originally stemming from Gupta and George (2016), and find areas for building an overall *big data analytics* capability. An example of this could be that an organization with investments in tangible resources like hardware, software and access to *big data* has a low data-driven culture or lack of technical skills. By using this construct as a benchmark, the CIO could identify these weak resources and take necessary actions.

Furthermore, we find that *big data analytics capabilities* have a positive correlation with both *dynamic capabilities* and *competitive performance*. This should help organizations understand that, depending on usage, they could increase their *operational capabilities* in both marketing capability and technology capability as in, for instance, market exploitation and efficient production/services. They could also increase their sensing capability so they become better at identifying opportunities and to exploit these.

We also find that *dynamic capabilities* has a positive correlation with *competitive performance*. Since most companies in this age have more agile capabilities, they also have a higher competitive advantage than those that lack those capabilities.

6.5 Limitations and future works

There were some limitations to our research. Although the different layouts of *big data analytics capabilities*, *dynamic capabilities*, *operational capabilities* and *competitive performance* (as well as environmental factors) had support from previous research, the composition of the research model, as a whole, is complex. We did not achieve good values that add significance to H4, which is the association between *operational capabilities* and *competitive performance*. This was unexpected since previous research has confirmed that this association has a positive correlating effect. We also did not have significant values on any of our environmental factors. Although this also has been partly researched earlier (in simplified models that address either *dynamic capabilities* → *competitive performance* or *operational* *capabilities* → *competitive performance*), we found no significance. Fine tuning measurements and increasing the sample from the population may lead to more significant values.

We achieve good and reliable data in the answers to our survey. The respondents that represented the organizations were mostly CIOs or *big data analytics* managers. All organizations used *big data analytics* solutions. A limitation we see are the questions in the survey. They could have better value if divided between several respondents within each organization. One respondent could answer questions related the CIOs position and another (CEO?) could answer more of the business side questions. This might remove personal misunderstandings or misconceptions related to personnel without the insight into the necessary organizational information. Kim et al. (2011) did this in their survey and achieved answers with less bias. This is of course difficult in our limited population (or at least the ones that are using *big data analytics* solutions today).

Overall, research based on self-reported questionnaires has its own limitations. For instance, participants may have varying degrees of understanding of knowledge or interpretations of particular questions. For instance, participants might assess the state of change (e.g. ED1, ED2) in different ways. What might be considered "quick" for some of them, can also be considered not so "quick" by others. Furthermore, people interpret and use scales differently. What one might rate as seven (totally agree) might be voted with six by someone else that also agrees with the statement. Some people like to use the extreme values of scales to position themselves while others avoid the extremes.

An important factor when looking at *competitive performance* is the time lag factor. Investments into capabilities might not provide results immediately but will need time to grow into *competitive performance*. We did some limited testing on this factor and did not see any big differences in the correlations. Though, they were a little smaller in values. This test was conducted on 75 respondents which is below the sample threshold. Increasing the sample i.e., getting more responses could provide enough data to identify differences. Also, while *big data analytics capabilities* are maturing in organizations and other organizations start using *big data analytics* solutions, possibilities for more participations increases and researchers could more easily include the time factor in their future research.

More responses might also provide possibilities to obtain results by looking at certain industries or different organization size. We did look mostly for big organizations and it would be interesting to see if small organizations also have the same results.

It would be interesting to measure which part of *big data analytics capabilities* has most correlation with *dynamic capabilities* or *operational capabilities*. This is outside our scope but seems interesting as the usage of *big data analytics* may possibly lead to different combinations of *big data analytics capabilities*. This could save costs and increase maximum exploitation of organizational resources.

In our survey, we had questions related to information governance and innovative capabilities. These were not included in the model or this master thesis except in the Appendix 6. We think those are important elements that belong in the model but since we did not include these in our literature review, we admit that we lack enough knowledge and need further understanding.

Another research study may provide this insight. The data collected is still valuable as it can be used by subsequent research studies.

To sum-up what we think future research should address:

1. Refine our model, indicators and sample of responders.
2. Try and find significance between *operational capabilities* and *competitive performance*.
3. Try and use multiple respondents per organization.
4. Time-lag should be accounted for. The same applies to industry and organization size.
5. More insight behind *big data analytics capabilities* and the connection to *dynamic capabilities* or *operational capabilities*.
6. Include information governance and innovative capabilities.

These, in addition to further improvements of the *big data analytics capabilities* construct could improve the understanding of *big data analytics capabilities* and its contribution to *competitive performance.*

7. Conclusion

The aim of this study has been to shed light on companies' use of *big data analytics* and how it leads to *competitive performance* by answering the following research question:

> *"Through what paths are big data analytic capabilities transformed into competitive performance "*

The research question was answered based on survey data from IT managers in 107 companies in the Nordic region. The data was analysed using partial least squares structural equation modelling (PLS-SEM).

Prior to the analysis, we conducted a systematic literature review and an explorative case study to better our understanding, further shape our research agenda and form a conceptual model. The model included two paths that we assumed could explain how *big data analytics* capability led to *competitive performance*. These were *dynamic capabilities* and *operational capabilities*.

The analysis showed significant support for three out of the four postulated hypotheses. There appeared to be a strong positive correlation between the company's grade of *big data analytics capability* towards *dynamic capability* and *operational capability*. Which meant that companies that were better composed to exploit the combination of data (*big data*) and IT components generally would better at: 1) sensing and seizing opportunities and reconfiguring their resource base accordingly and 2) improve and streamlining the day-to-day operations. Firstly, this may be due to the fact that companies, by proper use of *big data analytics*, get a better insight and a better basis for action in moving environments. Secondly, the use of *big data analytics* is closely related to the use of autonomous systems and artificial intelligence that holds huge potential to streamline and automate work processes.

There was also a medium positive correlation between *dynamic capability* and *competitive performance.* This means that companies that would better align and configure their resources according to market changes, generally also would have better performance than their competitors. Companies with a better ability to sense and transform their way of living according to market movements are more likely to have acquired business solutions that yield higher performance than those who have not.

However, we did not find any significant correlation between the concept of *operational capability* and *competitive performance*. The reason for this might be that the questions were not precise enough.

In response to the research question we propose, we can thus say that *big data analytics capabilities* lead to *competitive performance* through *dynamic capabilities* which can be seen as a mediator.

Technology leaders who want to invest in *big data analytics* should see the concept as more than just a technology concept and should develop distinctive capabilities to better handle the ocean of data.

Epilogue

The sun where shining and the birds were singing when our two researchers put their final touches to their master thesis before submitting the report to be judged. They were confident they had provided as good as they could manage and they were proud of their results. Indeed, the article they were introduced to in 2015 (Constantiou & Kallinikos, 2015) turned out to be a blessing in disguise and it made them start working towards what later became a master's thesis about the *big data* phenomenon. Now that the paper is finally delivered, they could turn their attention to their surroundings again and perhaps get some much-needed sleep. Or so they thought... First, they had to present and defend the thesis...

8. References

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R. & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics, 182*, 113-131. doi:10.1016/j.ijpe.2016.08.018

Ambrosini, V. & Bowman, C. (2009). What are dynamic capabilities and are they a useful construct in strategic management? *International journal of management reviews, 11*(1), 29-49. doi:10.1111/j.1468-2370.2008.00251.x

Andersson, O. (2012). *Experiment!: planning, implementing and interpreting*. West Sussex, UK: John Wiley & Sons.

Association for Information Systems. (2011). Senior Scholars' Basket of Journals. Retrieved from https://aisnet.org/?SeniorScholarBasket

Baller, S., Dutta, S. & Lanvin, B. (2016). *Innovating in the Digital Economy The Global Information Technology Report 2016* Retrieved from http://www3.weforum.org/docs/GITR2016/WEF_GITR_Full_Report.pdf

Bhimani, A. (2015). Exploring big data's strategic consequences. *Journal of Information Technology, 30*(1), 66-69. doi:10.1057/jit.2014.29

Cenfetelli, R. T. & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS quarterly, 33*(4), 689-707.

Chakravarty, A., Grewal, R. & Sambamurthy, V. (2013). Information technology competencies, organizational agility, and firm performance: Enabling and facilitating roles. *Information Systems Research, 24*(4), 976-997. doi:10.1287/isre.2013.0500

Chang, Y. B. & Gurbaxani, V. (2013). An empirical analysis of technical efficiency: The role of it intensity and competition. *Information Systems Research, 24*(3), 561-578. doi:10.1287/isre.1120.0438

Chen, D. Q., Preston, D. S. & Swink, M. (2015). How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *Journal of Management Information Systems, 32*(4), 4-39. doi:10.1080/07421222.2015.1138364

Chen, H., Chiang, R. H. & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly, 36*(4), 1165-1188.

Chen, Y., Wang, Y., Nevo, S., Jin, J., Wang, L. & Chow, W. S. (2014). IT capability and organizational performance: the roles of business process agility and environmental factors. *European Journal of Information Systems, 23*(3), 326-342. doi:10.1057/ejis.2013.4

Chi, L., Ravichandran, T. & Andrevski, G. (2010). Information technology, network structure, and competitive action. *Information Systems Research, 21*(3), 543-570. doi:10.1287/isre.1100.0296

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (Vol. 295, pp. 295-336). New Jersey: Lawrence erlbaum associates.

Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment, 7*(3), 309-319. doi:10.1037/1040-3590.7.3.309

Clarke, R. (2016). Big data, big risks. *Information Systems Journal, 26*(1), 77-90. doi:10.1111/isj.12088

Constantiou, I. D. & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology, 30*(1), 44-57. doi:10.1057/jit.2014.17

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: Sage publications.
Dess, G. G. & Beard, D. W. (1984). Dimensions of organizational task environments. *Administrative science quarterly, 29*(1), 52-73. doi:10.2307/2393080
Drnevich, P. L. & Kriauciunas, A. P. (2011). Clarifying the conditions and limits of the contributions of ordinary and dynamic capabilities to relative firm performance. *Strategic management journal, 32*(3), 254-279. doi:10.1002/smj.882
Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods, 4*(2), 144-192.
El Sawy, O. A. & Pavlou, P. A. (2008). IT-enabled business capabilities for turbulent environments. *MIS Quarterly Executive, 7*(3), 139-150.
Emani, C. K., Cullot, N. & Nicolle, C. (2015). Understandable big data: A survey. *Computer science review, 17*, 70-81. doi:10.1016/j.cosrev.2015.05.002
Espinosa, J. A. & Armour, F. (2016). *The Big Data Analytics Gold Rush: A Research Framework for Coordination and Governance.* Paper presented at the System Sciences (HICSS), 2016 49th Hawaii International Conference, Koloa, HI, USA.
European Commision. (2012). What is an SME? Retrieved from http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en
Fainshmidt, S., Pezeshkan, A., Frazier, M. L., Nair, A. & Markowski, E. (2016). Dynamic Capabilities and Organizational Performance: A Meta-Analytic Evaluation and Extension. *Journal of Management Studies, 53*(8), 1348-1380. doi:10.1111/joms.12213
Ferrier, W. J., Holsapple, C. W. & Sabherwal, R. (2010). Editorial commentary—Digital systems and competition. *Information Systems Research, 21*(3), 413-422. doi:10.1287/isre.1100.0304
Finn.no. (2017). Jobbsøk. Retrieved from https://www.finn.no/job/browse.html
Finnson, I. (2017, 24.01.2017). Smartere Analytics – Effektivitet og lønnsomhet gjennom Big Data. Retrieved from http://www.thecognizant.no/events/smartere-analytics-effektivitet-og-lonnsomhet-gjennom-big-data/
Forbes. (2017). The World's Biggest Public Companies. Retrieved from https://www.forbes.com/global2000/list/
Fornell, C. & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of marketing research, 18*(3), 382-388.
GalbRaith, J. R. (2014). Organizational design challenges resulting from big data. *Journal of Organization Design, 3*(1), 2-13.
Garmaki, M., Boughzala, I. & Wamba, S. F. (2016). *The effect of big data analytics capability on firm performance*. Paper presented at the PACIS 2016 Proceedings, Chiayi, Taiwan.
Garson, G. D. (2016). *Partial Least Squares: Regression & Structural Equation Models*. Asheboro, USA: Statistical Publishing Associates.
Gaskin, J. (2012). Gaskination's Statistics. Retrieved from http://youtube.com/Gaskination
George, G., Haas, M. R. & Pentland, A. (2014). Big data and management. *Academy of Management Journal, 57*(2), 321-326. doi:10.5465/amj.2014.4002
Gold, A. H. & Arvind Malhotra, A. H. S. (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems, 18*(1), 185-214.
Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic management journal, 17*(S2), 109-122.

Grant, R. M. (2016). *Contemporary strategy analysis: Text and cases edition*. Padstow, Cornwall, Great Britain: John Wiley & Sons.
Gupta, M. & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management, 53*(8), 1049-1064.
Hair, J. F., Ringle, C. M. & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice, 19*(2), 139-152.
Hair Jr, J. F., Hult, G. T. M., Ringle, C. & Sarstedt, M. (2013). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks, California: Sage Publications.
Helfat, C. E., Finkelstein, S., Mitchell, W., Peteraf, M., Singh, H., Teece, D. & Winter, S. G. (2009). *Dynamic capabilities: Understanding strategic change in organizations*. 350 Main Street, Malden, USA: John Wiley & Sons.
Hellevik, O. (2011). *Forskningsmetode i sosiologi og statsvitenskap*. Oslo: Universitetsforlaget.
Henseler, J., Ringle, C. M. & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135.
Henseler, J., Ringle, C. M. & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing *New challenges to international marketing* (pp. 277-319). Bingley, UK: Emerald Group Publishing Limited.
Indeed. (2017). Jobbsøk. Retrieved from https://dk.indeed.com/jobs
Israel, M. & Hay, I. (2006). *Research ethics for social scientists*. London: Sage.
Janssen, M., van der Voort, H. & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research, 70*, 338-345. doi:10.1016/j.jbusres.2016.08.007
Jobbsafari. (2017). Jobbsøk. Retrieved from https://www.jobbsafari.se/jobbsoekning
Kapital 500. (2016). Kapital 500 største. Retrieved from http://www.kapital500.no/
Kim, G., Shin, B., Kim, K. K. & Lee, H. G. (2011). IT capabilities, process-oriented dynamic capabilities, and firm financial performance. *Journal of the Association for Information Systems, 12*(7), 487-517.
Kirkpatrick, M. (2010, 4 aug, 2010). Google CEO Schmidt: "People Aren't Ready for the Technology Revolution". Retrieved from http://readwrite.com/2010/08/04/google_ceo_schmidt_people_arent_ready_for_the_tech/
Kline, R. (2011). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.
Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford publications.
Kumar, R. L. & Stylianou, A. C. (2014). A process model for analyzing and managing flexibility in information systems. *European Journal of Information Systems, 23*(2), 151-184. doi:10.1057/ejis.2012.53
Kwon, O., Lee, N. & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management, 34*(3), 387-394.
LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S. & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review, 52*(2), 21-32.
Li, D.-y. & Liu, J. (2014). Dynamic capabilities, environmental dynamism, and competitive advantage: Evidence from China. *Journal of Business Research, 67*(1), 2793-2799. doi:10.1016/j.jbusres.2012.08.007

Li, S., Shang, J. & Slaughter, S. A. (2010). Why do software firms fail? Capabilities, competitive actions, and firm survival in the software industry from 1995 to 2007. *Information Systems Research, 21*(3), 631-654. doi:10.1287/isre.1100.0281

Lim, J.-H., Stratopoulos, T. C. & Wirjanto, T. S. (2011). Path dependence of dynamic information technology capability: An empirical investigation. *Journal of Management Information Systems, 28*(3), 45-84. doi:10.2753/MIS0742-1222280302

Lind, D., Marchal, W. & Wathen, S. (2007). *Basic Statistics for Business and Economics with Student CD*: McGraw-Hill/Irwin.

Liu, H., Ke, W., Wei, K. K. & Hua, Z. (2013). The impact of IT capabilities on firm performance: The mediating roles of absorptive capacity and supply chain agility. *Decision Support Systems, 54*(3), 1452-1462. doi:10.1016/j.dss.2012.12.016

Lowry, P. B. & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE Transactions on Professional Communication, 57*(2), 123-146.

Lu, Y. & Ramamurthy, K. (2010). Proactive or reactive IT leaders? A test of two competing hypotheses of IT innovation and environment alignment. *European Journal of Information Systems, 19*(5), 601-618. doi:10.1057/ejis.2010.36

Mata, F. J., Fuerst, W. L. & Barney, J. B. (1995). Information technology and sustained competitive advantage: A resource-based analysis. *MIS quarterly, 19*(4), 487-505. doi:10.2307/249630

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. & Barton, D. (2012). Big data. *The management revolution. Harvard Bus Rev, 90*(10), 61-67.

Meetup. (2017). Largest Big Data Meetups. Retrieved from https://www.meetup.com/topics/big-data/

Midtbø, T. (2013). *Regresjonsanalyse for samfunnsvitere: med eksempler i SPSS* (4th ed.). Oslo: Universitetsforlaget.

Mikalef, P., Framnes, V. A., Danielsen, F., Krogstie, J. & Olsen, D. H. (2017). *Big Data Analytics Capability: Antecedents and Business Value*. Paper presented at the PACIS 2017, Langkawi.

Mikalef, P. & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research, 70*, 1-16. doi:10.1016/j.jbusres.2016.09.004

Newkirk, H. E. & Lederer, A. L. (2006). The effectiveness of strategic information systems planning under environmental uncertainty. *Information & Management, 43*(4), 481-501. doi:10.1016/j.im.2005.12.001

Nunally, J. C. & Bernstein, I. H. (1978). *Psychometric theory*. New York: McGraw-Hill.

Oates, B. J. (2006). *Researching information systems and computing*. London: Sage.

Ortbach, K., Plattfaut, R., Poppelbuß, J. & Niehaves, B. (2012). *A dynamic capability-based framework for business process management: Theorizing and empirical application.* Paper presented at the System Sciences (HICSS), 2012 45th Hawaii International Conference, Maui, HI, USA.

PACIS 2017. (2017). Pacific-Asia Conference on Information Systems 2017. Retrieved from http://www.pacis2017.org/

Pavlou, P. A. & El Sawy, O. A. (2006). From IT leveraging competence to competitive advantage in turbulent environments: The case of new product development. *Information Systems Research, 17*(3), 198-227. doi:10.1287/isre.1060.0094

Pavlou, P. A. & El Sawy, O. A. (2011). Understanding the elusive black box of dynamic capabilities. *Decision Sciences, 42*(1), 239-273. doi:10.1111/j.1540-5915.2010.00287.x

Petter, S., Straub, D. & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS quarterly, 31*(4), 623-656.

Protogerou, A., Caloghirou, Y. & Lioukas, S. (2012). Dynamic capabilities and their indirect impact on firm performance. *Industrial and Corporate Change, 21*(3), 615-647. doi:10.1093/icc/dtr049

Quaadgras, A., Ross, J. & Beath, C. (2013). You May Not Need Big Data After All. *Harvard business review, 91*, 90-98.

Rai, A., Patnayakuni, R. & Seth, N. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS quarterly, 30*(2), 225-246.

Rai, A. & Tang, X. (2010). Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios. *Information Systems Research, 21*(3), 516-542. doi:10.1287/isre.1100.0299

Ringle, C. M., Wende, S. & Becker, J.-M. (2015). SmartPLS 3. Retrieved from https://www.smartpls.com/

Roberts, N., Campbell, D. E. & Vijayasarathy, L. R. (2016). Using Information Systems to Sense Opportunities for Innovation: Integrating Postadoptive Use Behaviors with the Dynamic Managerial Capability Perspective. *Journal of Management Information Systems, 33*(1), 45-69. doi:10.1080/07421222.2016.1172452

RocketReach LLC. (2017). RocketReach. Retrieved from https://rocketreach.co/

Ross, J. W., Beath, C. M. & Quaadgras, A. (2013). You may not need big data after all. *Harvard business review, 91*(12), 90-+.

Roth, A. V. & Jackson III, W. E. (1995). Strategic determinants of service quality and performance: Evidence from the banking industry. *Management science, 41*(11), 1720-1733. doi:10.1287/mnsc.41.11.1720

Russom, P. (2011). Big data analytics. *TDWI Best Practices Report, Fourth Quarter, 19*(4), 1-34.

Sarmadawy, H. (2017). Dette er løsepengeviruset «WannaCry». Retrieved from http://www.vg.no/nyheter/innenriks/data-og-nett/dette-er-loesepengeviruset-wannacry/a/23997644/

Schilke, O. (2014a). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. *Strategic management journal, 35*(2), 179-203. doi:10.1002/smj.2099

Schilke, O. (2014b). Second-order dynamic capabilities: How do they matter? *The Academy of Management Perspectives, 28*(4), 368-380. doi:10.5465/amp.2013.0093

Schryen, G. (2013). Revisiting IS business value research: what we already know, what we still need to know, and how we can get there. *European Journal of Information Systems, 22*(2), 139-169.

Sharma, R., Mithas, S. & Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems, 23*(4), 433-441. doi:10.1057/ejis.2014.17

Slavec, A. & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review for Central and South-Eastern Europe, 14*(1), 39-62.

SurveyGizmo. (2017). Professional Survey Solution. Retrieved from https://www.surveygizmo.com/

Tallon, P. P., Ramirez, R. V. & Short, J. E. (2013). The information artifact in IT governance: toward a theory of information governance. *Journal of Management Information Systems, 30*(3), 141-178.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic management journal, 28*(13), 1319-1350.
Teece, D. J. (2014). The foundations of enterprise performance: Dynamic and ordinary capabilities in an (economic) theory of firms. *The Academy of Management Perspectives, 28*(4), 328-352. doi:10.5465/amp.2013.0116
Teece, D. J. (2015). Intangible assets and a theory of heterogeneous firms. In A. Bounfour & T. Miyagawa (Eds.), *Intangibles, Market Failure and Innovation Performance* (pp. 217-239). Switzerland: Springer.
Teece, D. J., Pisano, G. & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal, 18*(7), 509-533.
Teo, T. S., Srivastava, S. C. & Jiang, L. (2008). Trust and electronic government success: An empirical study. *Journal of Management Information Systems, 25*(3), 99-132.
UCLA. (2017). FAQ: What are the differences between one-tailed and two-tailed tests? Retrieved from https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-differences-between-one-tailed-and-two-tailed-tests/
Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R. & Cleven, A. (2009). *Reconstructing the giant: On the importance of rigour in documenting the literature search process*. Paper presented at the ECIS, Verona, Italy.
Wade, M. (2014). Dynamic capabilities. Retrieved from https://is.theorizeit.org/wiki/Dynamic_capabilities
Wamba, S. F., Akter, S., Edwards, A., Chopin, G. & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165*, 234-246. doi:10.1016/j.ijpe.2014.12.031
Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R. & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356-365. doi:10.1016/j.jbusres.2016.08.009
Wang, Y., Kung, L., Ting, C. & Byrd, T. A. (2015). *Beyond a technical perspective: understanding big data capabilities in health care.* Paper presented at the System Sciences (HICSS), 2015 48th Hawaii International Conference, Kaua'i, Hawaii.
Webster, J. & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly, 26*(2), xiii-xxiii.
Werner, S., Brouthers, L. E. & Brouthers, K. D. (1996). International risk and perceived environmental uncertainty: the dimensionality and internal consistency of Miller's measure. *Journal of International Business Studies, 27*(3), 571-587. doi:10.1057/palgrave.jibs.8490144
Wheeler, B. C. (2002). NEBIC: A dynamic capabilities theory for assessing net-enablement. *Information Systems Research, 13*(2), 125-146. doi:10.1287/isre.13.2.125.89
Wilden, R. & Gudergan, S. P. (2015). The impact of dynamic capabilities on operational marketing and technological capabilities: investigating the role of environmental turbulence. *Journal of the Academy of Marketing Science, 43*(2), 181-199. doi:10.1007/s11747-014-0380-y
Wilden, R., Gudergan, S. P., Nielsen, B. B. & Lings, I. (2013). Dynamic capabilities and performance: strategy, structure and environment. *Long Range Planning, 46*(1), 72-96. doi:10.1016/j.lrp.2012.12.001
Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic management journal, 24*(10), 991-995. doi:10.1002/smj.318
Woerner, S. & Wixom, B. H. (2015). Big data: extending the business strategy toolbox. *Journal of Information Technology, 30*(1), 60-62. doi:10.1057/jit.2014.31

Wolfswinkel, J. F., Furtmueller, E. & Wilderom, C. P. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems, 22*(1), 45-55.
Wong, K. K.-K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin, 24*(1), 1-32.
Wu, S. J., Melnyk, S. A. & Flynn, B. B. (2010). Operational capabilities: The secret ingredient. *Decision Sciences, 41*(4), 721-754. doi:10.1111/j.1540-5915.2010.00294.x
Wu, X., Zhu, X., Wu, G.-Q. & Ding, W. (2014). Data mining with big data. *IEEE transactions on knowledge and data engineering, 26*(1), 97-107. doi:10.1109/TKDE.2013.109
Zachariassen, S. (2017). Telenor advarer: Løsepengevirus på ferde i Norge. Retrieved from https://www.abcnyheter.no/penger/2017/02/23/195280146/telenor-advarer-losepengevirus-pa-ferde-i-norge
Zahra, S. A. & Garvis, D. M. (2000). International corporate entrepreneurship and firm performance: The moderating effect of international environmental hostility. *Journal of business venturing, 15*(5), 469-492. doi:10.1016/S0883-9026(99)00036-1
Zahra, S. A. & George, G. (2002). The net-enabled business innovation cycle and the evolution of dynamic capabilities. *Information Systems Research, 13*(2), 147-150. doi:10.1287/isre.13.2.147.90
Zhao, J. L., Fan, S. & Hu, D. (2014). Business challenges and research directions of management analytics in the big data era. *Journal of Management Analytics, 1*(3), 169-174. doi:10.1080/23270012.2014.968643

9. Appendix

Appendix 1: Articles included in the literature review

Table 26: Articles from Literature review (part A)

Literature Review - part A

Research question: "Which assets constitutes firm’s ability to use 'big data' solutions?"

		Search phrases									
Author (Year)	**Title**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
European Journal of Information Systems											
Sharma et al., 2014	Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations			X		X	X	X			X
Information Systems Journal											
Clarke, 2016	Big data, big risks					X		X			X
Journal of Information Technology											
Woerner & Wixom, 2015	Big data: extending the business strategy toolbox - Commentary					X		X			
Bhimani, 2015	Exploring big data's strategic consequences					X		X			X
Constantiou & Kallinikos, 2015	New games, new rules: big data and the changing context of strategy										X
Journal of MIS											
D.Q. Chen et al., 2015	How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management					X	X		X		X
Tallon et al., 2013	The Information Artifact in IT Governance: Toward a Theory of Information Governance					X	X	X	X		X
MIS Quarterly											
Chen et al., 2012	Business Intelligence and Analytics: From Big Data to Big Impact						X				X
Gupta & George, 2016	Toward the development of a big data analytics capability	From supervisors									
Wamba et al., 2017	Big data analytics and firm performance: Effects of dynamic capabilities	From supervisors									
Wamba et al., 2015	How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal casestudy	From supervisors									
Kwon et al., 2014	Data quality management, data usage experiance and acquisition intention of big data analyticts	From supervisors									
Galbraith, 2014	Organizational Design Challenges Resulting from Big Data	From supervisors									
Emani et al., 2015	Understandable Big Data: A survey	From supervisors									
Wu et al., 2014	Data Mining with Big Data	From supervisors									
Espinosa & Armour, 2016	The Big Data Analytics Gold Rush: A Research Framework for Coordination and Governance	From supervisors									
Janssen et al., 2016	Factors influencing big data decision-making quality	From supervisors									
Wang et al., 2015	Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care	Forward Backward searches									
Garmaki et al., 2016	The effect of big data analytics capability on firm performance	Forward Backward searches									
Akteret al., 2016	How to improve firm performance using big data analytics capability and business strategy alignment?	Forward Backward searches									

Table 27: Articles from Literature review (part B)

Literature Review - part B

Research question: "How do big data lead to competitive advantages?"

		Search phrases								
Author (Year)	**Title**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
European Journal of Information Systems										
Sharma et al., 2014	Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations	X			X	X				
Kumar & Stylianou, 2014	A process model for analyzing and managing flexibility in information systems					X				
Lu & Ramamurthy, 2010	Proactive or reactive IT leaders? A test of two competing hypotheses of IT innovation and environment alignment						X			
Information Systems Research										
Li et al., 2010	Why do software firms fail? Capabilities, competitive actions, and firm survival in the software industry from 1995 to 2007				X					
Chi et al., 2010	Information technology, network structure, and competitive action				X	X				
Chakravarty et al., 2013	Information technology competencies, organizational agility, and firm performance: Enabling and facilitating roles						X			
Ferrier et al., 2010	Editorial Commentary-Digital Systems and Competition						X			
Rai & Tang, 2010	Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios						X			
Wheeler, 2002	NEBIC: A dynamic capabilities theory for assessing net-enablement							X		
Zahra & George, 2002	The net-enabled business innovation cycle and the evolution of dynamic capabilities							X		
Journal of AIS										
Kim et al., 2011	IT capabilities, process-oriented dynamic capabilities, and firm financial performance				X	X		X		
Journal of MIS										
Chen et al., 2015	How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management	X			X	X				
Roberts et al., 2016	Using Information Systems to Sense Opportunities for Innovation: Integrating Postadoptive Use Behaviors with the Dynamic Managerial Capability Perspective	X			X	X		X		
Lim et al., 2011	Path dependence of dynamic information technology capability: An empirical investigation						X	X		
MIS Quarterly										
Sawy & Pavlou, 2008	IT-enabled business capabilities for turbulent environments				X					

Drnevich & Kriauciunas, 2011	Clarifying the conditions and limits of the contributions of ordinary and dynamic capabilities to relative firm performance	From supervisors
Fainshmidt et al., 2016	Dynamic Capabilities and Organizational Performance: A Meta-Analytic Evaluation and Extension	From supervisors
Pavlou & Sawy, 2011	Understanding the Elusive Black Box of Dynamic Capabilities	From supervisors
Pavlou & Sawy, 2006	From IT Leveraging Competence to Competitive Advantage in Turbulent Environments: The Case of New Product Development	From supervisors
Protogerou et al., 2012	Dynamic capabilities and their indirect impact on firm performance	From supervisors
Wilden et al., 2013	Dynamic Capabilities and Performance: Strategy, Structure and Environment	From supervisors
Li & Liu, 2014	Dynamic capabilities, environmental dynamism, and competitive advantage: Evidence from China	From supervisors
Schilke, 2014a	Second-order dynamic capabilities: How do they matter?	From supervisors
Schilke, 2014b	On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of	From supervisors
Teece et al., 1997	Dynamic Capabilities and Strategic Management	From supervisors
Teece, 2007	Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance	From supervisors
Winter, 2003	Understanding dynamic capabilities	From supervisors
Liu et al., 2013	The impact of IT capabilities on firm performance: The mediating roles of absorptive capacity and supply chain agility	From supervisors
Mikalef & Pateli, 2017	Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA	From supervisors
Wamba et al., 2017	Big data analytics and firm performance: Effects of dynamic capabilities	From supervisors
Wilden & Gudergan, 2015	The impact of dynamic capabilities on operational marketing and technological capabilities: investigating the role	From supervisors
Wu et al., 2010	Operational Capabilities: The Secret Ingredient	From supervisors

Appendix 2: Concept matrix - Literature review

Table 28: Concept matrix from literature review 1/2 (part A)

		Research method					Type of paper				Location						Industry		
	Article / Concept	Quantitative study	Qualitative study	Case study	Theoretical study	Literature review	Article	Commentaries	Editorial	Conference paper	Europe	North America	South America	Oceania	Africa	Asia	E-government	Health	E-commerce
1	Gupta & George, 2016	X					X					X							
2	Clarke, 2016				X		X							X					
3	Sharma et al., 2014				X				X			X		X		X			
4	Chen et al., 2012				X	X	X					X							
5	Wamba et al., 2017	X					X									X			X
6	D. Q. Chen et al., 2015	X			X		X					X							
7	Wamba et al., 2015		X	X		X	X							X				X	
8	Constantiou & Kallinikos, 2015				X		X				X								
9	Kwon et al., 2014	X					X									X			
10	Tallon et al., 2013		X				X					X							
11	Galbraith, 2014				X		X				X								
12	Garmaki et al., 2016				X					X	X								
13	Emani et al., 2015		X				X												
14	Wu et al., 2014				X		X					X		X		X			
15	Wang et al., 2015				X					X		X						X	
16	Akter et al., 2016	X				X	X					X							
17	Espinosa & Armour, 2016				X					X									
18	Janssen et al., 2016			X			X				X						X		
19	Woerner & Wixom, 2015				X			X				X							
20	Bhimani, 2015				X			X			X								

Table 29: Concept matrix from literature review 2/2 (part A)

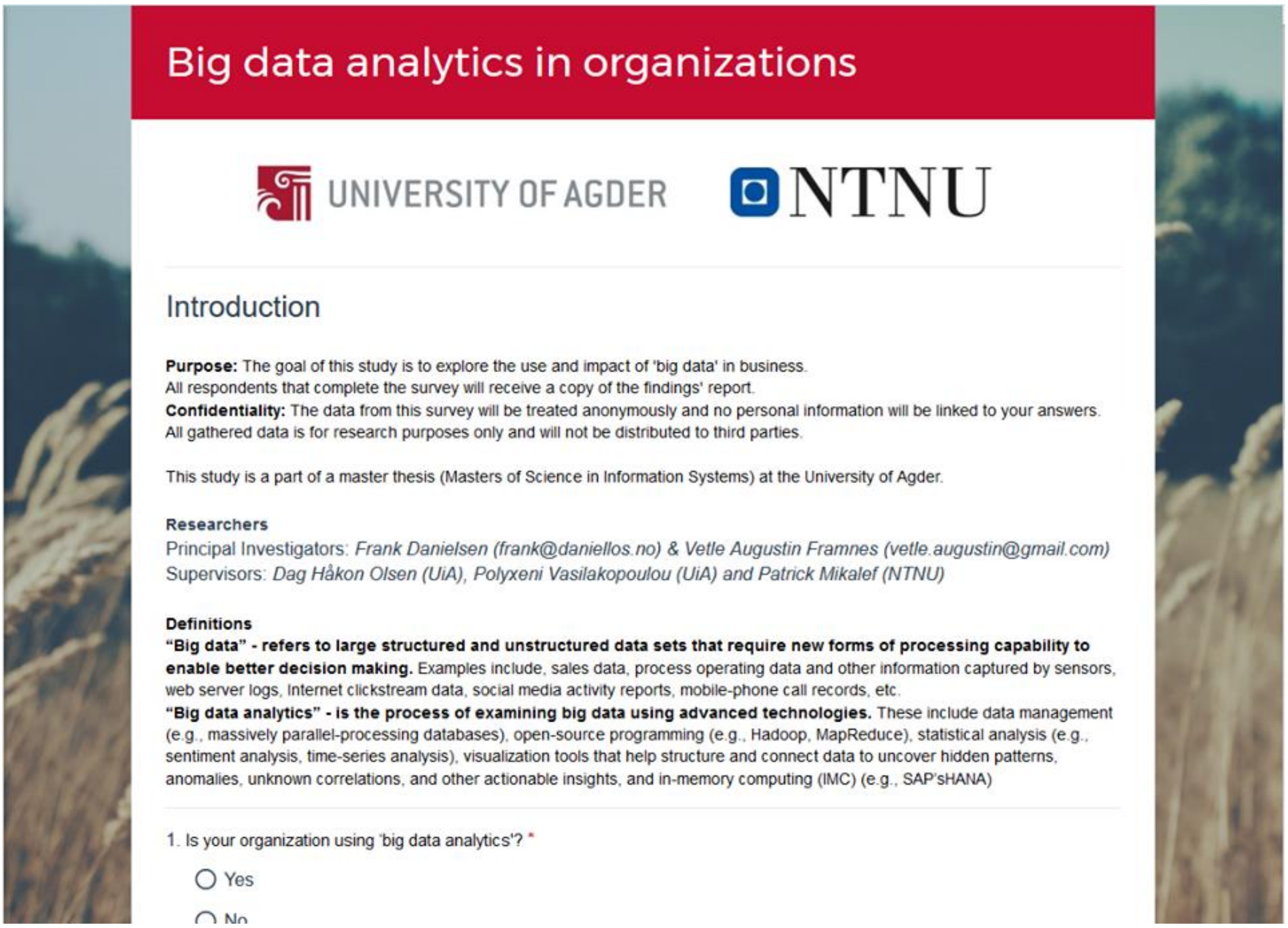
	Article / Concept	Big Data Analytics Capability															
		Personnel expertise Capability				Management Capability					Infrastructure Capability			Organizational capability			Presentation capability
		Technical knowledge	Technology management capability	Business Knowledge	Relational knowledge	Planning	Investment	Coordination	Control	Top Management support	Connectivity	Compatibility	Modularity	Data-driven culture	Agility	Analytical capability	Visualization
1	Gupta & George, 2016	X		X			X		X*		X	X		X			
2	Clarke, 2016		X						X*								X
3	Sharma et al., 2014							X	X*					X			
4	Chen et al., 2012	X		X			X							X			
5	Wamba et al., 2017	X	X	X	X	X	X	X	X		X	X	X				
6	D. Q. Chen et al., 2015	X					X			X		X		X			
7	Wamba et al., 2015	X		X			X	X	X*	X	X	X		X	X	X	
8	Constantiou & Kallinikos, 2015											X					
9	Kwon et al., 2014								X*					X			
10	Tallon et al., 2013								X*	X				X			
11	Galbraith, 2014									X	X			X	X		
12	Garmaki et al., 2016	X							X*	X	X	X		X	X		
13	Emani et al., 2015	X							X*		X		X				X
14	Wu et al., 2014			X					X*								
15	Wang et al., 2015				X			X	X*					X		X	X
16	Akter et al., 2016	X	X	X	X	X	X	X	X		X	X	X				
17	Espinosa & Armour, 2016							X	X								
18	Janssen et al., 2016	X			X			X	X								
19	Woerner & Wixom, 2015														X		
20	Bhimani, 2015													X	X		

* Data Governance

Table 30: Concept matrix from literature review (part B)

		Research method				Type of paper				Location						Concept			
	Article / Concept	Quantitative study	Qualitative study	Theoretical study	Literature review	Article	Commentary	Editorial	Conference paper	Europe	North America	South America	Oceania	Africa	Asia	Dynamic capability	Operational capability	(Competitive) performance	Environmental turbulence
1	Sharma et al., 2014			X		X							X			X		X	
2	Kumar & Stylianou, 2014			X		X					X							X	X
3	Lu & Ramamurthy, 2010		X			X					X					X		X	X
4	Li et al., 2010	X				X					X						X	X	X
5	Chi et al., 2010	X				X					X					X		X	
6	Ferrier et al., 2010			X		X	X	X			X					X		X	X
7	Rai & Tang, 2010	X				X					X							X	X
8	Wheeler, 2002			X		X					X					X			
9	Zahra & George, 2002			X		X	(X)				X					X		X	
10	Kim et al., 2011	X				X									X	X	X	X	
11	Chen et al. 2015	X				X					X					X	X	X	X
12	Roberts et al. 2016	X				X					X					X		X	
13	Chakravarty et al., 2013	X				X				X								X	X
14	Lim et al., 2011	X				X					X					X		X	
15	El Sawy & Pavlou, 2008			X		X					X					X	X		X
16	Drnevich & Kriauciunas, 2011	X				X					X					X	X	X	
17	Fainshmidt et al., 2016	X				X					X					X		X	
18	Pavlou & El Sawy, 2011	X				X					X					X	X	X	X
19	Pavlou & El Sawy, 2006	X				X					X					X		X	X
20	Protogerou et al., 2012	X				X				X						X	X	X	
21	Wilden, et al. 2013	X				X							X			X		X	
22	Li & Liu, 2014	X				X									X	X		X	X
24	Schilke, 2014a	X				X					X					X		X	X
23	Schilke 2014b	X				X					X					X		X	
25	Teece et al., 1997			X		X					X					X		X	X
26	Teece, 2007			X		X					X					X	X	X	X
27	Winter, 2003			X		X					X					X	X		
28	Liu et al., 2013	X				X									X	X		X	
29	Mikalef & Pateli, 2017	X				X				X						X		X	X
30	Wamba et al., 2017	X				X							X			X		X	
31	Wilden & Gudergan, 2015	X				X							X			X	X	X	X
32	Wu et al., 2010	X				X					X					X	X	X	

Appendix 3: Image excerpt of the survey



Big data analytics in organizations

UNIVERSITY OF AGDER NTNU

Introduction

Purpose: The goal of this study is to explore the use and impact of 'big data' in business.
All respondents that complete the survey will receive a copy of the findings' report.
Confidentiality: The data from this survey will be treated anonymously and no personal information will be linked to your answers.
All gathered data is for research purposes only and will not be distributed to third parties.

This study is a part of a master thesis (Masters of Science in Information Systems) at the University of Agder.

Researchers
Principal Investigators: *Frank Danielsen (frank@daniellos.no) & Vetle Augustin Framnes (vetle.augustin@gmail.com)*
Supervisors: *Dag Håkon Olsen (UiA), Polyxeni Vasilakopoulou (UiA) and Patrick Mikalef (NTNU)*

Definitions
"Big data" - refers to large structured and unstructured data sets that require new forms of processing capability to enable better decision making. Examples include, sales data, process operating data and other information captured by sensors, web server logs, Internet clickstream data, social media activity reports, mobile-phone call records, etc.
"Big data analytics" - is the process of examining big data using advanced technologies. These include data management (e.g., massively parallel-processing databases), open-source programming (e.g., Hadoop, MapReduce), statistical analysis (e.g., sentiment analysis, time-series analysis), visualization tools that help structure and connect data to uncover hidden patterns, anomalies, unknown correlations, and other actionable insights, and in-memory computing (IMC) (e.g., SAP'sHANA)

1. Is your organization using 'big data analytics'? *

○ Yes

○ No

Figure 12: Excerpt from page one of the survey

Appendix 4: Research model

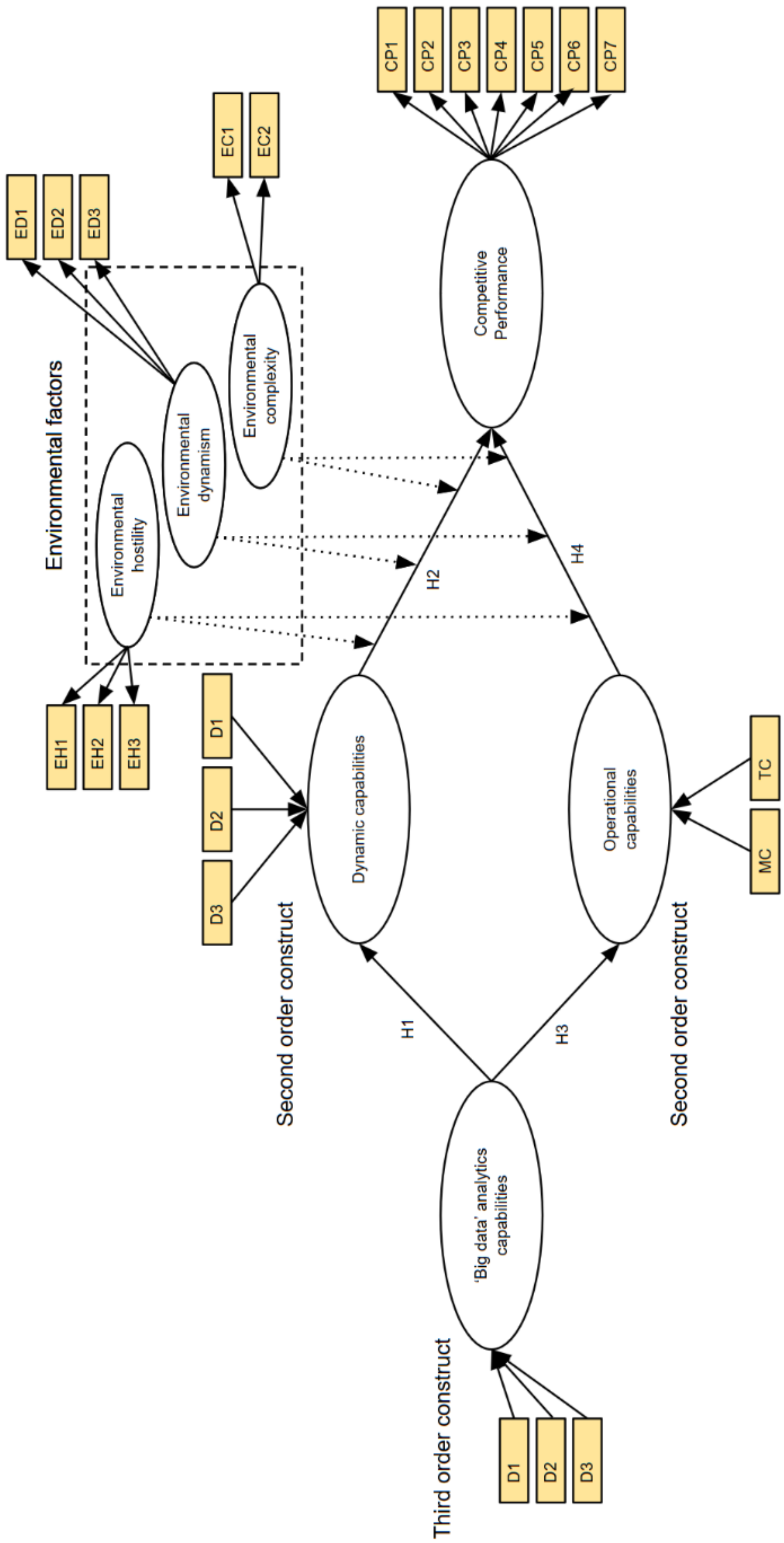


Figure 13: Research model, after two-stage reduction

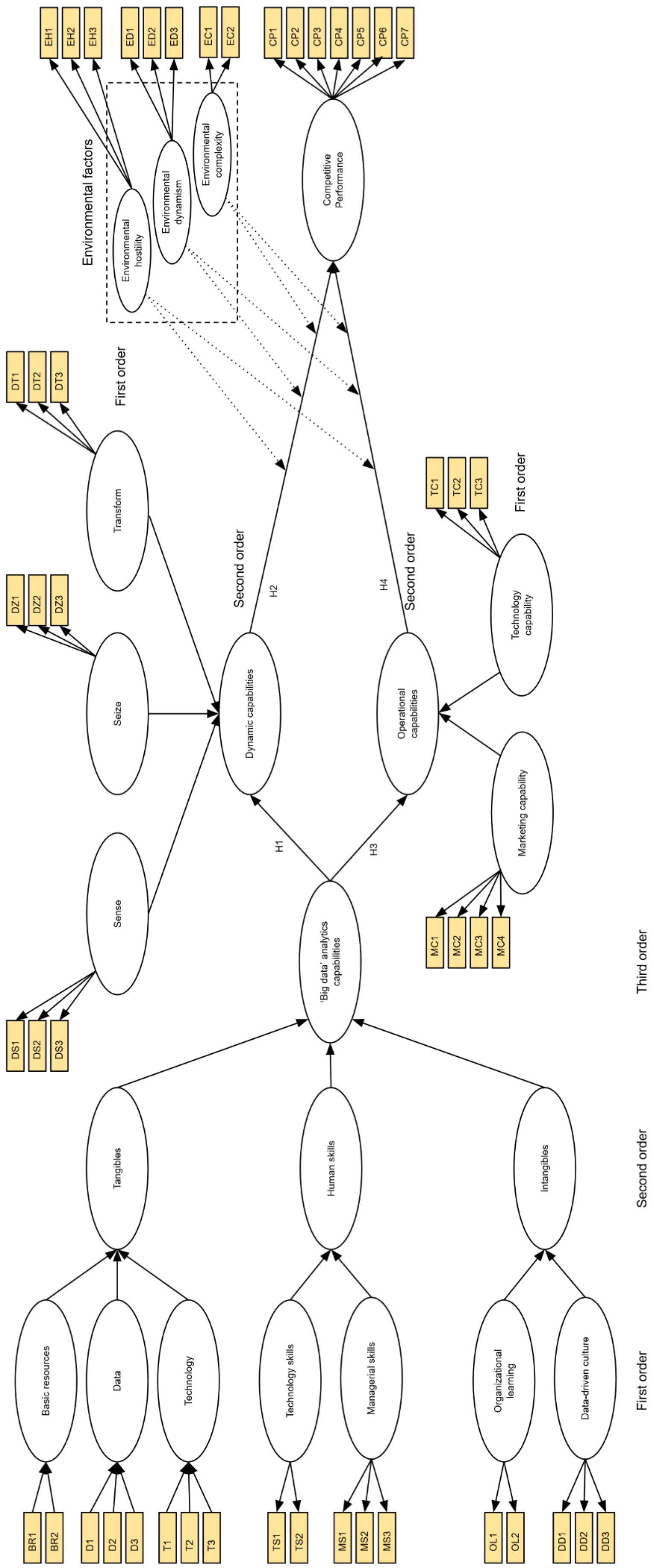


Figure 14: Research model, full version

Appendix 5: Reliability and validity definitions

t-value: Used to establish the significance of the formative relations. Can be calculated to find the p-values. Smart PLS 3 also has functions for displaying the corresponding p-values. Both research sites and textbooks like Hellevik (2011, p. 403) and Andersson (2012, Appendix B) provides tables that help quickly identify p-values.

p-value: Used to establish the significance of the formative relations. It is measured as a a number with 3 decimals (for instance 0.05). The number represents the percentage calculation for the chance that an erroneous rejection of a true null hypothesis would occur (Hellevik, 2011, p. 390).

Variance inflation factor (VIF): Quantifies the severity of collinearity among the indicators in a formative measurement model. The VIF is directly related to the tolerance value (Hair Jr et al., 2013, pp. 124-125, 165).

Adequacy coefficient (R2a): Used to evaluate the validity of the items of formative constructs (Edwards, 2001).

Composite reliability: a measure of internal consistency reliability, which, unlike Cronbach's alpha, does not assume equal indicator loadings. Smart PLS 3 measures this through PLS Algorithm. The value should be above 0.708 (in exploratory research, 0.60 to 0.70 is considered acceptable) (Hair Jr et al., 2013, p. 115).

Cronbach's alpha: A measure of internal consistency reliability that assumes equal indicator loadings. Smart PLS 3 measures this through PLS Algorithm. The values should be above 0.708. In the context of PLSSEM, composite reliability is considered a more suitable criterion of reliability. However, Cronbach's alpha still represents a conservative measure of internal consistency reliability (Hair Jr et al., 2013, p. 115).

Indicator reliability: is the square of a standardized indicator's outer loading. It represents how much of the variation in an item is explained by the construct and is referred to as the variance extracted from the item (Hair Jr et al., 2013, p. 115). Also, loadings below 0.70 was considered removed from the construct (Gupta & George, 2016).

Average variance extracted (AVE): is a measure of convergent validity. It is the degree to which a latent construct explains the variance of its indicators. The AVE values should be above the minimum level of 0.50 (Hair Jr et al., 2013, pp. 110, 115).

Cross loadings: is an overview over an indicator's correlation with other constructs in the model. This is to ensure that the indicator measures what it is intended to measure (Hair Jr et al., 2013, p. 115). This table also show the indicators loadings and therefore also establish that all values are above the threshold for loadings (0.708).

Fornell-Larcker criterion: is a measure of discriminant validity that compares the square root of each construct's average variance extracted with its correlations with all other constructs in the model (Hair Jr et al., 2013, pp. 105-107).

Heterotrait-monotrait ratio (HTMT): a way to assess discriminant validity. According to Henseler et al. (2015), this is a better alternative than Fornell-Larker criterion and cross-loadings in assessing discriminant validity. This function has been implemented in SmartPLS 3 and can be viewed by running a PLS calculation.

Path coefficients: are estimated path relationships in the structural model (i.e., between the constructs in the model). They correspond to standardized betas in a regression analysis (Hair Jr et al., 2013, p. 116).

Coefficient of determination (R2): is a measure of the proportion of an endogenous construct's variance that is explained by its predictor constructs (Hair Jr et al., 2013, p. 115). The values can be interpreted as substantial (0.75), moderate (0.5) and weak (0.25) (Hair et al., 2011).

Predictive relevance (Q2): is a measure of predictive relevance based on the blindfolding technique (Hair Jr et al., 2013, pp. 202, 203).

Effect size (f2): is a measure used to assess the relative impact of a predictor construct on an endogenous construct (Hair Jr et al., 2013, p. 201).

Effect size (q2): is a measure used to assess the relative predictive relevance of a predictor construct on an endogenous construct (Hair Jr et al., 2013, p. 203).

Appendix 6: Survey instrument

Table 31: Survey questions, mean and standard devision

Name	Formative/ Reflective	Question	Mean	S.Dev.
Background question				
BG0	Control	Is your organization using 'big data analytics'?	-	-
BG1	Control	When did your organization start using 'big data analytics' solutions? *(measured in years)*	3.18	1.50
BG2	Control	How old is your organization? (measured in years)	4.27	0.86
BG3	Control	Please indicate the size-class of your organization. (Number of employees)	3.59	0.70
BG4	Control	In which industry does your organization operate? (multiple choice + textbox)	-	-
Environmental complexity				
EC1	Reflective	customer buying habits	4.75	1.79
EC2	Reflective	nature of competition	4.51	1.59
Environmental hostility				
EH1	Reflective	price competition	4.86	1.90
EH2	Reflective	competition in product/service quality	4.80	1.73
EH3	Reflective	competition in product/service differentiation	4.94	1.68
Environmental dynamism				
ED1	Reflective	In our industry, products and services become obsolete quickly	2.98	1.69
ED2	Reflective	The product/services technologies in our industry change quickly	4.29	1.81
ED3	Reflective	Our competitors' behaviours exhibit a lot of variability	4.04	1.60
Basic resources				
BR1	Formative	adequately funded	4.67	1.52
BR2	Formative	given enough time to achieve their objectives	4.41	1.38
Technology skills				
TS1	Reflective	has the right skills to accomplish their jobs successfully	4.83	1.37
TS2	Reflective	is well trained	4.72	1.36
Technology				

T1	Formative	parallel computing approaches (e.g., Hadoop) to big data processing	4.15	2.28
T2	Formative	different data visualization tools	5.42	1.49
T3	Formative	new forms of databases such as Not Only SQL(NoSQL) for storing data	4.58	2.21
Data-driven culture				
DD1	Reflective	We base our decisions on data rather than on instinct	4.71	1.34
DD2	Reflective	We are willing to override our own intuition when data contradict our viewpoints	4.88	1.41
DD3	Reflective	We continuously coach our employees to make decisions based on data	4.45	1.40
Data				
D1	Formative	We have access to very large, unstructured, or fast-moving data for analysis	5.28	1.47
D2	Formative	We integrate data from multiple internal sources into a data warehouse or mart for easy access	5.50	1.46
D3	Formative	We integrate external data with internal to facilitate high-value analysis of our business environment	4.41	1.73
Organizational learning				
OL1	Reflective	We are able to acquire new and relevant knowledge	5.19	1.08
OL2	Reflective	We have made concerted efforts for the exploitation of existing competencies and exploration of new knowledge	4.75	1.39
Managerial skills				
MS1	Reflective	understand the business need of (and collaborate with) other functional managers, suppliers, and customers to determine opportunities that big data might bring to our business.	4.85	1.34
MS2	Reflective	coordinate big data-related activities in ways that support other functional managers, suppliers, and customers	4.60	1.30
MS3	Reflective	understand and evaluate the output extracted from big data	5.04	1.11
Marketing capabilities and Technological capabilities				
MC1	Reflective	Market knowledge	5.63	1.13
MC2	Reflective	Control and access to distribution channels	5.26	1.23
MC3	Reflective	Advantageous relationships with customers	5.25	1.27

MC4	Reflective	Established customer base	5.67	1.33
TC1	Reflective	Efficient and effective production/services	5.07	1.14
TC2	Reflective	Economies of scales and technical expertise	5.27	1.25
TC3	Reflective	Technological capabilities and equipment	5.45	1.24
Sensing				
DS1	Reflective	We frequently scan the environment to identify new business opportunities	5.10	1.41
DS2	Reflective	We often review our product development efforts to ensure they are in line with what the customers want	5.11	1.42
DS3	Reflective	We use established processes to identify target market segments, changing customer needs and customer innovation	4.88	1.47
Seizing				
DZ1	Reflective	drafting various potential solutions	4.75	1.29
DZ2	Reflective	evaluating and selecting potential solutions	4.83	1.23
DZ3	Reflective	starting on a detailed plan to carry out a potential solution	4.77	1.32
Transforming				
DT1	Reflective	Create new or substantially changed ways of achieving our targets and objectives	4.63	1.34
DT2	Reflective	Adjusting our business processes in response to shifts in our business priorities	4.58	1.44
DT3	Reflective	Reconfiguring our business processes in order to come up with new productive assets	4.49	1.42
Competitive performance				
CP1	Reflective	profitability	4.76	1.47
CP2	Reflective	return on investment (ROI)	4.78	1.30
CP3	Reflective	growth in market share	4.66	1.55
CP4	Reflective	sales growth	4.67	1.47
CP5	Reflective	rapid response to market demand	4.47	1.44
CP6	Reflective	in reducing operating costs	4.47	1.32
CP7	Reflective	increasing customer satisfaction	4.84	1.24
The following questions (indicators) were distributed to the respondents but not included in this thesis				
Information governance capabilities (Procedural)				
IGP1	Reflective	setting retention policies (e.g. time to live) of data	4.50	1.33
IGP2	Reflective	backup routines	5.80	1.14

IGP3	Reflective	establishing/monitoring access (e.g. user access) to data	5.33	1.26
IGP4	Reflective	classifying data according to value	4.25	1.47
IGP5	Reflective	monitoring costs versus value of data	3.73	1.56
Information governance capabilities (Structural and relational)				
IGS1	Reflective	have identified key IT and non-IT decision makers to have the responsibility regarding data ownership, value analysis and cost management.	4.36	1.56
IGS2	Reflective	use steering committees to oversee and assess data values and costs	3.76	1.59
IGR1	Reflective	educate users and non-IT managers regarding storage utilization and costs	3.47	1.52
IGR2	Reflective	develop communications regarding policy effectiveness and user needs	4.00	1.43
Innovation capabilities (incremental and radical)				
INI1	Reflective	Innovations that reinforce our prevailing product/service lines	5.01	1.14
INI2	Reflective	Innovations that reinforce our existing expertise in prevailing products/services	4.96	1.10
INR1	Reflective	Innovations that make our prevailing product/service lines obsolete	3.91	1.46
INR2	Reflective	Innovations that make our existing expertise in prevailing products/services obsolete	3.85	1.38